

Handbook of  
Accelerator Physics  
and Engineering

Second Edition

**This page intentionally left blank**

Handbook of  
Accelerator Physics  
and Engineering

Second Edition

*edited by*

**Alexander Wu Chao**

*SLAC National Accelerator Laboratory, USA*

**Karl Hubert Mess**

*CERN, Switzerland*

**Maury Tigner**

*Cornell University, USA*

**Frank Zimmermann**

*CERN, Switzerland*



NEW JERSEY • LONDON • SINGAPORE • BEIJING • SHANGHAI • HONG KONG • TAIPEI • CHENNAI

*Published by*

World Scientific Publishing Co. Pte. Ltd.

5 Toh Tuck Link, Singapore 596224

*USA office:* 27 Warren Street, Suite 401-402, Hackensack, NJ 07601

*UK office:* 57 Shelton Street, Covent Garden, London WC2H 9HE

**British Library Cataloguing-in-Publication Data**

A catalogue record for this book is available from the British Library.

**HANDBOOK OF ACCELERATOR PHYSICS AND ENGINEERING**

**Second Edition**

Copyright © 2013 by World Scientific Publishing Co. Pte. Ltd.

*All rights reserved. This book, or parts thereof, may not be reproduced in any form or by any means, electronic or mechanical, including photocopying, recording or any information storage and retrieval system now known or to be invented, without written permission from the Publisher.*

For photocopying of material in this volume, please pay a copying fee through the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, USA. In this case permission to photocopy is not required from the publisher.

ISBN 978-981-4415-84-2

ISBN 978-981-4417-17-4 (pbk)

# Preface

This Handbook is a product of the world community of accelerator physicists and engineers. The first edition was issued September, 1998. With the continued advancing of the accelerator field, a second edition appears now in order.

This is not a textbook but rather a collection of information useful to professionals in research, design, construction, and operation of accelerators. The Handbook has been prepared by more than 200 experienced experts from across the spectrum of accelerator related institutions and to them great thanks are due.

In addition to content, a high priority has been given to portability of the book. This has led to a sacrifice of some aesthetics in order to make the text as compact as possible. For that, our apologies go to users and authors alike.

Singularly important are the references to be found at the end of each subsection. Here the user will find locations of tutorial material as well as reliable detail for further reading. The references for the most part are not intended to be exhaustive or to indicate priority of discovery or invention, but rather to provide a reliable lead into the literature. In addition, a detailed index gives access to occurrences of important subjects and concepts to be found herein.

The fees and royalties that would normally be paid to authors and editors are donated to provide scholarships to the CERN Accelerator School and to the US Particle Accelerator School.

Authors and Editors have made great efforts to find and eliminate errors. Nevertheless we recognize that there will be errors and have provided for errata to appear on a Handbook website:

<http://www.worldscientific.com/worldscibooks/10.1142/8543>

Please help in this community effort by sending suggestions for corrections by e-mail to the address [achao@slac.stanford.edu](mailto:achao@slac.stanford.edu).

## Editors

Alexander Wu Chao, Stanford, California  
Maury Tigner, Ithaca, New York  
Frank Zimmermann, Geneva, Switzerland  
Karl-Hubert Meß, Sülfeld, Germany

*July 2012*

## Acknowledgments to Second Edition

The editors hereby acknowledge with appreciation the support of the Managements of SLAC National Accelerator Laboratory at Stanford University, the Cornell Laboratory for Accelerator-based Sciences and Education, and the European Organization for Nuclear Research, CERN, without which this work would have been impossible. We also wish to give our special thanks to Jeanne Butler and World Scientific Publishing Company staff for countless essential technical help.

## Acknowledgments to First Edition

The editors hereby acknowledge with deepest appreciation and thank the support of the Directors of SLAC at Stanford University and of the Laboratory of Nuclear Studies at Cornell University without which this work would have simply been impossible. We also wish to give special thanks to Scott Berg, Indiana University, and Tom Hays, Cornell, for essential help with the technical complexities of typesetting software and to Jim Wahl, SLAC and Don Miller, Cornell for their indispensable help with the illustrations. For continual guidance and help with the intricacies of computer hardware and software, great thanks are due to Tom Knight, SLAC.

# Table of Contents

|  |          |
|--|----------|
| <b>Preface</b>   | <b>v</b> |
| <b>1 INTRODUCTION</b>  | <b>1</b> |
| 1.1 HOW TO USE THIS BOOK . . . . .   | 1        |
| 1.2 NOMENCLATURE . . . . .   | 1        |
| 1.3 FUNDAMENTAL CONSTANTS . . . . .  | 3        |
| 1.4 UNITS AND CONVERSIONS . . . . .  | 4        |
| 1.4.1 Units <i>A.W. Chao</i> . . . . .   | 4        |
| 1.4.2 Conversions <i>M. Tigner</i> . . . . .   | 4        |
| 1.5 FUNDAMENTAL FORMULAE <i>A.W. Chao</i> . . . . .                                  | 5        |
| 1.5.1 Special Functions . . . . .  | 5        |
| 1.5.2 Curvilinear Coordinate Systems . . . . .                                       | 6        |
| 1.5.3 Electromagnetism . . . . .   | 6        |
| 1.5.4 Kinematical Relations . . . . .  | 7        |
| 1.5.5 Vector Analysis . . . . .  | 8        |
| 1.5.6 Relativity . . . . .   | 8        |
| 1.6 GLOSSARY OF ACCELERATOR TYPES . . . . .  | 8        |
| 1.6.1 Antiproton Sources <i>K. Gollwitzer, J. Marriner</i> . . . . .                 | 8        |
| 1.6.2 Betatron <i>M. Tigner</i> . . . . .  | 10       |
| 1.6.3 Colliders <i>D. Hartill</i> . . . . .  | 12       |
| 1.6.4 Cyclotron <i>H. Blosser, M. Craddock</i> . . . . .                             | 14       |
| 1.6.5 Electrostatic Accelerator <i>J. Ferry</i> . . . . .                            | 17       |
| 1.6.6 FFAG Accelerator <i>M.K. Craddock</i> . . . . .                                | 18       |
| 1.6.7 High Voltage Electrodynmaic Accelerators <i>M.R. Cleland</i> . . . . .         | 22       |
| 1.6.8 Induction Linac <i>R. Bangerter</i> . . . . .                                  | 24       |
| 1.6.9 Industrial Accelerators <i>R. Hellborg</i> . . . . .                           | 27       |
| 1.6.10 Laser, Wakefield and Plasma Accelerators <i>J. Rosenzweig</i> . . . . .       | 31       |
| 1.6.11 Linear Accelerators for Electrons <i>G.A. Loew</i> . . . . .                  | 34       |
| 1.6.12 Linear Accelerators for Protons <i>S. Henderson, A. Aleksandrov</i> . . . . . | 36       |
| 1.6.13 Livingston Chart <i>D. Hartill</i> . . . . .                                  | 40       |
| 1.6.14 Medical Applications of Accelerators <i>J. Alonso</i> . . . . .               | 40       |
| 1.6.14.1 Radiation therapy . . . . .   | 40       |
| 1.6.14.2 Radioisotopes . . . . .   | 43       |
| 1.6.15 $\mu^+\mu^-$ Collider <i>R. Palmer</i> . . . . .                              | 44       |
| 1.6.15.1 Collider . . . . .  | 44       |
| 1.6.15.2 Muon storage ring neutrino factories . . . . .                              | 46       |
| 1.6.15.3 Technical challenges . . . . .  | 46       |
| 1.6.16 Neutron Sources <i>J. Wei, H. Lengeler</i> . . . . .                          | 47       |
| 1.6.17 Pulsed High Voltage Devices <i>J.A. Nation, D.A. Hammer</i> . . . . .         | 51       |
| 1.6.18 Radio Frequency Quadrupole <i>J. Staples</i> . . . . .                        | 52       |
| 1.6.19 Rhodotron <i>Y. Jongen, M. Abs</i> . . . . .                                  | 55       |
| 1.6.20 Storage Rings <i>W. Fischer</i> . . . . .                                     | 56       |

|  |           |
|--|-----------|
| 1.6.21 Synchrotrons <i>C. Zhang, S.X. Fang</i> . . . . .   | 58        |
| 1.6.22 Two-Beam Accelerator <i>J.P. Delahaye</i> . . . . .   | 60        |
| 1.7 ACCELERATOR COMPUTER CODES <i>R. Ryne</i> . . . . .  | 62        |
| <b>2 BEAM DYNAMICS</b>   | <b>65</b> |
| 2.1 PHASE SPACE . . . . .  | 65        |
| 2.1.1 Linear Betatron Motion <i>D.A. Edwards, M.J. Syphers</i> . . . . .   | 65        |
| 2.1.2 Longitudinal Motion <i>D.A. Edwards, M.J. Syphers</i> . . . . .  | 66        |
| 2.1.3 Linear Coupled System <i>D.A. Edwards, M.J. Syphers</i> . . . . .  | 67        |
| 2.1.4 Orbital Eigen-Analysis for Electron Storage Ring <i>J.A. Ellison, H. Mais, G. Ripken</i> . . . . .                           | 68        |
| 2.2 OPTICS AND LATTICES . . . . .  | 71        |
| 2.2.1 Single Element Optics <i>K. Brown</i> . . . . .  | 71        |
| 2.2.2 3-D Multipole Expansion, Calculation of Transfer Maps from Field Data, Fringe Fields <i>M. Venturini, A. Dragt</i> . . . . . | 75        |
| 2.2.3 Lattices for Collider Storage Rings <i>E. Keil</i> . . . . .   | 77        |
| 2.2.4 Lattices for Low-Emissittance Light Sources <i>C. Steier</i> . . . . .   | 81        |
| 2.2.4.1 Lattice choices . . . . .  | 81        |
| 2.2.4.2 Chromaticity correction and nonlinear optimization . . . . .   | 83        |
| 2.2.4.3 Systematic lattice optimization techniques . . . . .   | 83        |
| 2.2.4.4 Evolution of light source lattices . . . . .   | 83        |
| 2.2.4.5 Ultimate storage rings . . . . .   | 84        |
| 2.2.5 Betatron Motion with Coupling of Two Degrees of Freedom <i>V. Lebedev, A. Burov</i> . . . . .                                | 85        |
| 2.3 NONLINEAR DYNAMICS . . . . .   | 89        |
| 2.3.1 Hamiltonian <i>K. Symon</i> . . . . .  | 89        |
| 2.3.1.1 General case . . . . .   | 89        |
| 2.3.1.2 Transverse motion . . . . .  | 91        |
| 2.3.1.3 Longitudinal motion . . . . .  | 91        |
| 2.3.1.4 Synchrobetatron coupling . . . . .   | 92        |
| 2.3.2 Tune Dependence on Momentum and Betatron Amplitudes<br><i>D.A. Edwards, M.J. Syphers</i> . . . . .                           | 93        |
| 2.3.3 Nonlinear Resonances <i>D.A. Edwards, M.J. Syphers</i> . . . . .   | 94        |
| 2.3.4 Synchrobetatron Resonances <i>A. Piwinski</i> . . . . .  | 94        |
| 2.3.5 Taylor Map, Hénon Map, Standard Map <i>A. Dragt</i> . . . . .  | 97        |
| 2.3.6 Lie Algebraic Methods <i>A. Dragt</i> . . . . .  | 99        |
| 2.3.7 Differential Algebraic Techniques <i>M. Berz</i> . . . . .   | 105       |
| 2.3.8 Symplectic Integration Methods <i>H. Yoshida</i> . . . . .   | 109       |
| 2.3.8.1 Methods of realization . . . . .   | 109       |
| 2.3.8.2 Symplectic method vs. nonsymplectic method . . . . .   | 111       |
| 2.3.9 Dynamic Aperture <i>A. Wolski</i> . . . . .  | 111       |
| 2.3.10 Decoherence <i>M.A. Furman</i> . . . . .  | 115       |
| 2.3.11 Momentum Compaction and Phase Slip Factor <i>K.Y. Ng</i> . . . . .  | 116       |
| 2.3.12 Nonlinear Dynamics Experiments <i>W. Fischer</i> . . . . .  | 118       |
| 2.3.13 Echo Effects <i>G.V. Stupakov</i> . . . . .   | 121       |
| 2.4 COLLECTIVE EFFECTS . . . . .   | 123       |
| 2.4.1 Collective Effects in High Energy Electron Linacs <i>K. Kubo, K. Yokoya, K. Thompson</i> . . . . .                           | 123       |
| 2.4.1.1 Single bunch longitudinal dynamics . . . . .   | 123       |
| 2.4.1.2 Multibunch longitudinal dynamics . . . . .   | 123       |
| 2.4.1.3 Single bunch transverse dynamics . . . . .   | 124       |
| 2.4.1.4 Multibunch transverse dynamics . . . . .   | 125       |
| 2.4.1.5 Effects of structure misalignment . . . . .  | 126       |
| 2.4.2 Collective Effects in Energy Recovery Linacs <i>G. Hoffstaetter</i> . . . . .  | 127       |

|          |  |     |
|----------|--|-----|
| 2.4.3    | Beam Loading <i>D. Boussard</i>  | 129 |
| 2.4.3.1  | Single-bunch passage in a cavity   | 129 |
| 2.4.3.2  | Cavity equivalent circuit  | 129 |
| 2.4.3.3  | Transmission of small modulations (AM and PM) through a cavity with beam loading               | 130 |
| 2.4.3.4  | Periodic beam loading at multiples of $f_0$  | 131 |
| 2.4.3.5  | Rf power needed for transient beam-loading correction  | 132 |
| 2.4.3.6  | Traveling-wave cavities  | 132 |
| 2.4.4    | Space-Charge Dominated Beams in Guns and Transport Lines <i>M. Ferrario</i>                    | 133 |
| 2.4.5    | Space Charge Effects in Circular Accelerators <i>B. Zotter</i>                                 | 137 |
| 2.4.5.1  | Direct space charge effects  | 137 |
| 2.4.5.2  | Betatron frequency shifts  | 138 |
| 2.4.6    | Beam Dynamics in Proton Linacs <i>S. Henderson, A. Aleksandrov</i>                             | 140 |
| 2.4.7    | Vlasov and Fokker-Planck Equations <i>B. Zotter</i>  | 144 |
| 2.4.8    | Potential Well Effect <i>B. Zotter</i>   | 145 |
| 2.4.9    | Single-Bunch Instabilities in Circular Accelerators <i>B. Zotter</i>                           | 147 |
| 2.4.10   | Sacherer Formulae <i>B. Zotter</i>   | 150 |
| 2.4.11   | Landau Damping <i>K.Y. Ng</i>  | 152 |
| 2.4.12   | Intrabeam Scattering and Touschek Effect <i>V. Lebedev</i>                                     | 155 |
| 2.4.13   | Ion Trapping, Beam-Ion Instabilities, and Dust <i>F. Zimmermann</i>                            | 159 |
| 2.4.13.1 | Ion trapping   | 159 |
| 2.4.13.2 | Dust particles   | 160 |
| 2.4.13.3 | Single-pass ion effects in storage rings and linacs  | 161 |
| 2.4.14   | Electron-Cloud Effect <i>M.A. Furman</i>   | 163 |
| 2.4.15   | Coherent Synchrotron Radiation Instability <i>G. Stupakov</i>                                  | 167 |
| 2.5      | <b>BEAM-BEAM EFFECTS</b>   | 169 |
| 2.5.1    | Beam-Beam Effects in Storage Ring Colliders <i>K. Hirata</i>                                   | 169 |
| 2.5.1.1  | Infinitely short bunches   | 169 |
| 2.5.1.2  | Long bunches   | 172 |
| 2.5.1.3  | Dispersion at IP, crossing angle   | 173 |
| 2.5.2    | Crab Waist Collision Scheme <i>M. Zobov</i>  | 174 |
| 2.5.3    | Beam-Beam Effects in Linear Colliders <i>P. Chen, D. Schulte</i>                               | 175 |
| 2.5.3.1  | Disruption   | 176 |
| 2.5.3.2  | Beamstrahlung  | 177 |
| 2.5.3.3  | Background and spent beam  | 178 |
| 2.5.4    | Parasitic Beam-Beam Effects and Separation Schemes <i>J.M. Jowett</i>                          | 180 |
| 2.5.4.1  | Separation schemes   | 180 |
| 2.5.4.2  | Long-range beam-beam effects   | 181 |
| 2.6      | <b>POLARIZATION</b>  | 183 |
| 2.6.1    | Thomas-BMT Equation <i>T. Roser</i>  | 183 |
| 2.6.2    | Spinor Algebra <i>T. Roser</i>   | 184 |
| 2.6.3    | Spin Rotators and Siberian Snakes <i>T. Roser</i>  | 185 |
| 2.6.4    | Depolarizing Resonances and Spin Flippers <i>T. Roser</i>                                      | 186 |
| 2.6.5    | Polarized Hadron Beams and Siberian Snakes <i>A.D. Krisch, M.A. Leonova</i>                    | 187 |
| 2.6.6    | Radiative Polarization in Electron Storage Rings <i>D.P. Barber, G. Ripken</i>                 | 191 |
| 2.6.7    | Computing Algorithms for $e^-/e^+$ Polarization in Storage Rings <i>D.P. Barber, G. Ripken</i> | 194 |
| 2.6.8    | Spin Matching in $e^-/e^+$ Rings <i>D.P. Barber, G. Ripken</i>                                 | 195 |
| 2.6.9    | Lie Algebra for Spin Motion <i>K. Yokoya</i>   | 198 |
| 2.7      | <b>BEAM COOLING</b>  | 199 |
| 2.7.1    | Stochastic Cooling <i>M. Blaskiewicz</i>   | 199 |
| 2.7.1.1  | Cooling rates  | 199 |

|          |  |            |
|----------|--|------------|
| 2.7.1.2  | Hardware . . . . .   | 201        |
| 2.7.1.3  | New techniques . . . . .   | 201        |
| 2.7.2    | Electron Cooling <i>S. Nagaitev</i> . . . . .  | 202        |
| 2.7.3    | Laser Cooling <i>J. Hangst, A. Noda</i> . . . . .  | 205        |
| 2.7.4    | Ionization Cooling <i>D. Neuffer</i> . . . . .   | 209        |
| 2.7.5    | Crystalline Beam <i>J. Wei</i> . . . . .   | 212        |
| <b>3</b> | <b>ELECTROMAGNETIC AND NUCLEAR INTERACTIONS</b>  | <b>215</b> |
| 3.1      | <b>SYNCHROTRON RADIATION</b> . . . . .   | 215        |
| 3.1.1    | Radiation of a Point Charge <i>H. Wiedemann</i> . . . . .  | 215        |
| 3.1.2    | Coherent Radiation <i>H. Wiedemann</i> . . . . .   | 216        |
| 3.1.3    | Bending Magnet Radiation <i>H. Wiedemann</i> . . . . .   | 217        |
| 3.1.4    | Synchrotron Radiation in Storage Rings <i>H. Wiedemann</i> . . . . .                             | 220        |
| 3.1.4.1  | Radiation integrals . . . . .  | 220        |
| 3.1.4.2  | Radiation damping . . . . .  | 220        |
| 3.1.4.3  | Quantum excitation . . . . .   | 221        |
| 3.1.4.4  | Equilibrium beam emittances . . . . .  | 221        |
| 3.1.4.5  | Damping wigglers . . . . .   | 222        |
| 3.1.4.6  | Quantum lifetimes . . . . .  | 222        |
| 3.1.5    | Undulator and Wiggler Radiation <i>H. Wiedemann</i> . . . . .                                    | 223        |
| 3.1.5.1  | Polarization of synchrotron radiation <i>H. Wiedemann</i> . . . . .                              | 225        |
| 3.1.6    | Transition and Diffraction Radiation <i>C. Thongbai</i> . . . . .                                | 226        |
| 3.1.7    | Coherent Synchrotron Radiation <i>S. Krinsky</i> . . . . .                                       | 227        |
| 3.1.8    | Free-Electron Lasers <i>Z. Huang, P. Schmüser</i> . . . . .                                      | 229        |
| 3.1.9    | Ultrashort X-ray Pulse Generation <i>A. Zholtens</i> . . . . .                                   | 233        |
| 3.1.10   | Compton/Thomson Sources <i>G.A. Krafft</i> . . . . .   | 235        |
| 3.1.10.1 | Luminosity description . . . . .   | 236        |
| 3.1.10.2 | Nonlinear scattering . . . . .   | 236        |
| 3.1.11   | “Short” Magnet and Edge Radiation <i>H. Wiedemann</i> . . . . .                                  | 237        |
| 3.1.12   | Beam Solid-Target Photon Physics <i>K. Ispiryan</i> . . . . .                                    | 237        |
| 3.2      | <b>IMPEDANCES AND WAKE FUNCTIONS</b> . . . . .   | 242        |
| 3.2.1    | Definitions and Properties of Impedances and Wake Functions <i>T. Suzuki</i> . . . . .           | 242        |
| 3.2.2    | Impedance Calculation, Frequency Domain <i>R.L. Gluckstern, S.S. Kurennoy</i> . . . . .          | 243        |
| 3.2.3    | Impedance Calculation, Time Domain <i>E. Gjonaj, T. Weiland</i> . . . . .                        | 248        |
| 3.2.4    | Explicit Expressions of Impedances and Wake Functions <i>K.Y. Ng, K. Bane</i> . . . . .          | 252        |
| 3.2.5    | Effective Impedance <i>T. Suzuki</i> . . . . .   | 262        |
| 3.2.6    | Parasitic Loss <i>P. Wilson, B. Zotter, Y.-H. Chin</i> . . . . .                                 | 263        |
| 3.2.7    | Trapped Modes <i>S.S. Kurennoy</i> . . . . .   | 265        |
| 3.3      | <b>PARTICLE-MATTER INTERACTION</b> . . . . .   | 266        |
| 3.3.1    | Basic Formulae <i>M. Tigner, A.W. Chao</i> . . . . .   | 266        |
| 3.3.2    | Beam and Luminosity Lifetime . . . . .   | 269        |
| 3.3.2.1  | Protons <i>N.V. Mokhov, V.I. Balbekov</i> . . . . .  | 269        |
| 3.3.2.2  | Electrons <i>M.S. Zisman</i> . . . . .   | 271        |
| 3.3.3    | Bhabha Scattering ( $e^+e^- \rightarrow e^+e^-$ ) <i>J.E. Spencer</i> . . . . .                  | 273        |
| 3.3.4    | Compton Scattering ( $e^\pm\gamma \rightarrow e^\pm\gamma$ ) <i>J.E. Spencer</i> . . . . .       | 275        |
| 3.3.5    | Limit of Focusing of Electron Beam due to Synchrotron Radiation <i>K. Oide</i> . . . . .         | 277        |
| 3.3.6    | Thermal Outgassing and Beam Induced Desorption <i>V. Baglin, E. Mahner, O. Gröbner</i> . . . . . | 278        |
| 3.3.7    | Photoemission and Secondary Emission <i>V. Baglin</i> . . . . .                                  | 283        |
| 3.3.7.1  | Photoemission . . . . .  | 283        |
| 3.3.7.2  | Secondary emission . . . . .   | 284        |
| 3.3.8    | Ionization Processes <i>F. Zimmermann</i> . . . . .  | 286        |

|          |   |            |
|----------|---|------------|
| 3.3.9    | Beam Induced Detector Backgrounds and Irradiation in $e^+e^-$ Colliders<br><i>S.D. Henderson, M. Sullivan</i> | 286        |
| 3.3.9.1  | Sources of detector backgrounds   | 287        |
| 3.3.9.2  | Detector and IR radiation tolerance and budget  | 290        |
| 3.3.9.3  | Detector background shielding   | 290        |
| 3.3.9.4  | Detector background and radiation estimation  | 291        |
| 3.3.10   | Particle Interactions and Beam-Induced Backgrounds and Radiation<br><i>N.V. Mokhov, S.I. Striganov</i>        | 293        |
| 3.3.11   | Beam Collimation <i>R. Assmann</i>  | 300        |
| 3.3.12   | Atomic and Nuclear Properties of Materials  | 308        |
| <b>4</b> | <b>OPERATIONAL CONSIDERATIONS</b>   | <b>311</b> |
| 4.1      | LUMINOSITY <i>M.A. Furman, M.S. Zisman</i>  | 311        |
| 4.2      | BRIGHTNESS <i>P. Elleaume, K.-J. Kim</i>  | 318        |
| 4.2.1    | Particle Beam   | 318        |
| 4.2.2    | Radiation Beam  | 319        |
| 4.2.2.1  | Introduction  | 319        |
| 4.2.2.2  | Bending magnet radiation  | 319        |
| 4.2.2.3  | Wiggler radiation   | 320        |
| 4.2.2.4  | Undulator radiation   | 320        |
| 4.2.2.5  | Brightness comparison   | 320        |
| 4.3      | LINAC OPERATION   | 321        |
| 4.3.1    | Operation of High Energy Electron Linacs <i>T.O. Raubenheimer</i>   | 321        |
| 4.3.2    | Operation of Superconducting Linacs <i>S. Schreiber</i>   | 326        |
| 4.3.3    | Halo in High Intensity Proton Linacs <i>T.O. Raubenheimer</i>   | 329        |
| 4.4      | RECIRCULATED ENERGY RECOVERY LINACS <i>S.A. Bogacz, D.R. Douglas, G.A. Krafft</i>                             | 329        |
| 4.4.1    | Recirculation and Energy Recovery   | 329        |
| 4.4.2    | System Architecture and Beam Dynamics Issues  | 331        |
| 4.5      | BUNCH COMPRESSION <i>P. Emma</i>  | 334        |
| 4.6      | OPERATION OF LINEAR-COLLIDER FINAL FOCUS SYSTEMS <i>A.A. Seryi, G.R. White</i>                                | 337        |
| 4.6.1    | Chromaticity Compensation   | 338        |
| 4.7      | OPERATION OF CIRCULAR ACCELERATORS  | 346        |
| 4.7.1    | Error Sources and Effects <i>D. Rice</i>  | 346        |
| 4.7.2    | Orbit and Lattice Function Measurements <i>D. Rice</i>  | 347        |
| 4.7.3    | Orbit Correction <i>J. Wenninger</i>  | 350        |
| 4.7.3.1  | Global orbit correction   | 350        |
| 4.7.3.2  | SVD algorithm   | 350        |
| 4.7.3.3  | MICADO algorithm  | 351        |
| 4.7.3.4  | Local orbit bumps   | 351        |
| 4.7.4    | Measurement and Diagnosis of Coupling and Solenoid Compensation <i>D. Rubin</i>                               | 352        |
| 4.7.4.1  | Sources of transverse coupling  | 352        |
| 4.7.4.2  | Solenoids   | 352        |
| 4.7.4.3  | Coupling matrix analysis  | 353        |
| 4.7.4.4  | Measurement of coupling   | 354        |
| 4.7.4.5  | Solenoid compensation   | 355        |
| 4.7.5    | Modeling and Control of Storage Rings Using Orbit Measurements <i>J. Safranek</i>                             | 357        |
| 4.7.6    | Emittance Dilution Effects <i>M. Syphers</i>  | 360        |
| 4.7.6.1  | Injection mismatch  | 360        |
| 4.7.6.2  | Diffusion processes   | 361        |

|          |   |            |
|----------|---|------------|
| 4.7.7    | Real-Time Measurement and Control of Tune, Coupling and Chromaticity<br><i>R. Jones, R. Steinhausen</i> | 362        |
| 4.7.7.1  | Tune measurement  | 362        |
| 4.7.7.2  | Chromaticity measurement  | 363        |
| 4.7.7.3  | Coupling measurement  | 363        |
| 4.7.7.4  | Real-time control of tune, coupling & chromaticity  | 364        |
| 4.7.8    | Measurement of Dispersion by Resonant Excitation <i>D.L.Rubin</i>                                       | 365        |
| 4.7.8.1  | Introduction  | 365        |
| 4.7.8.2  | Formalism   | 365        |
| 4.7.8.3  | Measurement of the coupling matrix  | 366        |
| 4.8      | TEMPORAL AND SPATIAL CORRELATIONS IN BPM MEASUREMENTS   | 367        |
| 4.8.1    | Non-Invasive Measurement for Linacs <i>J.Irwin, Y.Yan</i>   | 367        |
| 4.8.2    | Invasive Measurement for $e^+, e^-$ Circular Accelerators <i>J.Irwin, Y.Yan</i>                         | 370        |
| 4.8.3    | Virtual Models for Proton Circular Accelerators <i>Y.Yan</i>  | 373        |
| 4.9      | TRANSITION CROSSING <i>J.Wei</i>  | 373        |
| 4.10     | RF GYMNASTICS IN A SYNCHROTRON <i>R.Garoby</i>  | 376        |
| 4.10.1   | Adiabaticity  | 376        |
| 4.10.2   | Single Bunch Manipulations  | 376        |
| 4.10.3   | Multi-Bunch Manipulations   | 378        |
| 4.10.4   | Debunched Beam Manipulation   | 380        |
| 4.10.5   | Beam Manipulations with Broad-Band RF Systems   | 380        |
| 4.11     | ENERGY MEASUREMENT WITH ELECTRON BEAMS <i>J.Seeman</i>  | 381        |
| 4.12     | RING INJECTION AND EXTRACTION <i>G.H.Rees</i>   | 382        |
| 4.12.1   | Aspects of Slow Extraction <i>P.J.Bryant</i>  | 385        |
| 4.13     | COLLISION SCHEMES FOR RING COLLIDERS <i>M.Zobov, F.Zimmermann</i>                                       | 387        |
| 4.14     | BEAM-BEAM COMPENSATION SCHEMES <i>J.P.Koutchouk, V.Shiltsev</i>   | 391        |
| 4.14.1   | Compensation of the Head-On Beam-Beam Effect  | 391        |
| 4.14.2   | Compensation of the Long-Range Beam-Beam Effect   | 392        |
| 4.15     | SPACE CHARGE COMPENSATION (SCC) IN HADRON BEAMS <i>V.Shiltsev</i>                                       | 394        |
| 4.15.1   | Longitudinal SCC: Inductive Inserts   | 394        |
| 4.15.2   | Transverse SCC  | 395        |
| 4.16     | OPERATIONAL CONSIDERATIONS ON COOLING <i>S.Nagaitsev, M.Blaskiewicz</i>                                 | 396        |
| <b>5</b> | <b>MECHANICAL CONSIDERATIONS</b>  | <b>399</b> |
| 5.1      | MECHANICAL AND THERMAL PROPERTIES OF STRUCTURAL MATERIALS<br><i>M.Kuchnir</i>                           | 399        |
| 5.2      | MECHANICAL AND THERMAL PROPERTIES OF COMPOSITE SUPERCONDUCTORS <i>R.M.Scanlan</i>                       | 404        |
| 5.3      | THERMODYNAMIC & HYDRODYNAMIC PROPERTIES OF COOLANTS & CRYOGENS <i>M.McAshan</i>                         | 406        |
| 5.4      | CREEP AND STRESS RELAXATION IN ACCELERATOR COMPONENTS<br><i>F.Markley</i>                               | 410        |
| 5.5      | ELECTRIC AND MAGNETIC FORCES <i>M.Tigner</i>  | 411        |
| 5.6      | DEFLECTIONS AND BUCKLING <i>M.Tigner</i>  | 412        |
| 5.7      | PRACTICAL HEAT TRANSFER AND FLUID FLOW <i>M.McAshan, M.Tigner</i>                                       | 414        |
| 5.8      | FABRICATION OF NIOBIUM RF STRUCTURES <i>T.Hays, H.Padamsee, D.Proch</i>                                 | 420        |
| 5.9      | REFRIGERATION SYSTEMS <i>C.Rode, R.Ganni</i>  | 425        |
| 5.9.1    | Refrigerators   | 425        |
| 5.9.2    | Storage and Utilities   | 429        |
| 5.9.3    | Transfer Lines  | 429        |
| 5.10     | VACUUM SYSTEMS  | 431        |
| 5.10.1   | Requirements for Vacuum Systems <i>N.B.Mistry, Y.Li</i>   | 431        |

|          |   |            |
|----------|---|------------|
| 5.10.2   | Units, Conversions and Some Useful Formulae <i>N.B. Mistry, Y. Li</i>   | 432        |
| 5.10.3   | Conductance and Pressure Profiles <i>N.B. Mistry, Y. Li</i>   | 433        |
| 5.10.4   | Pumping Methods <i>N.B. Mistry, Y. Li</i>   | 435        |
| 5.10.5   | Instrumentation <i>N.B. Mistry, Y. Li</i>   | 439        |
| 5.10.6   | Vacuum Chamber Design and Fabrication <i>N.B. Mistry, Y. Li</i>   | 440        |
| 5.10.7   | Special Components in the Vacuum System <i>N.B. Mistry, Y. Li</i>   | 445        |
| 5.10.8   | Ceramic Vacuum Chamber Design <i>H.L. Phillips</i>  | 447        |
| 5.11     | ALIGNMENT <i>R. Ruland</i>  | 448        |
| 5.12     | MAGNET SUPPORTS AND ALIGNMENT <i>G. Bowden</i>  | 451        |
| 5.13     | GROUND VIBRATION <i>C. Montag, J. Rossbach</i>  | 453        |
| 5.13.1   | Basics  | 453        |
| 5.13.2   | Measurements  | 454        |
| 5.13.3   | Instruments   | 454        |
| 5.13.4   | Linacs  | 455        |
| 5.13.5   | Circular Accelerators   | 456        |
| 5.13.6   | Numerical Modeling  | 458        |
| 5.14     | VIBRATION CONTROL IN ACCELERATORS <i>C. Montag</i>  | 459        |
| 5.14.1   | General Considerations  | 459        |
| 5.14.2   | Passive Damping   | 459        |
| 5.14.3   | Active Stabilization  | 460        |
| <b>6</b> | <b>ELECTRICAL CONSIDERATIONS</b>  | <b>463</b> |
| 6.1      | PROPERTIES OF DIELECTRICS <i>M. Tigner</i>  | 463        |
| 6.2      | PROPERTIES OF CONDUCTORS, NORMAL AND SUPERCONDUCTING<br><i>R.M. Scanlan</i>                                       | 465        |
| 6.3      | PROPERTIES OF FERROMAGNETIC MATERIALS <i>M.Tigner</i>   | 471        |
| 6.4      | PERMANENT MAGNET MATERIALS <i>R.D. Schlueter</i>  | 471        |
| 6.5      | PROPERTIES OF LOSSY MATERIALS <i>E. Chojnacki</i>   | 473        |
| 6.6      | COMMON TRANSMISSION LINES AND CAVITIES <i>M. Tigner</i>   | 475        |
| 6.7      | RF PULSE COMPRESSION  | 481        |
| 6.7.1    | Passive Pulse Compression <i>Z.D. Farkas</i>  | 481        |
| 6.7.2    | Active Pulse Compression <i>S. Tantawi</i>  | 485        |
| 6.7.3    | Ultra-High-Power Multimoded Rf Components <i>S. Tantawi, C. Nantista</i>  | 486        |
| 6.8      | RF WINDOWS AND CAVITY COUPLING <i>R.M. Sundelin, H.L. Phillips</i>  | 491        |
| 6.9      | MULTIPACTING <i>D. Proch, J. Sekutowicz</i>   | 494        |
| 6.10     | POLYPHASE POWER CIRCUITS <i>M. Tigner</i>   | 496        |
| 6.11     | HIGH PRECISION POWER CONVERTERS <i>F. Bordry, J.P. Burne,<br/>M. Cerqueira Bastos</i>                             | 498        |
| 6.11.1   | Introduction to Magnet Power Converters   | 498        |
| 6.11.2   | Main Parameters of Magnet Power Converters  | 498        |
| 6.11.3   | Power Converter Topologies  | 498        |
| 6.11.3.1 | Thyristor controlled rectifier  | 498        |
| 6.11.3.2 | Switch-mode power converter   | 499        |
| 6.12     | HIGH ACCURACY IN POWER CONVERTERS FOR PARTICLE ACCELERATORS<br><i>F. Bordry, J.P. Burnet, M. Cerqueira Bastos</i> | 502        |
| 6.12.1   | Power Converter Control   | 502        |
| 6.12.2   | Current Measurement in Particle Accelerators  | 503        |
| 6.13     | HIGH-GRADIENT LIMITATIONS IN ROOM TEMPERATURE RF LINACS<br><i>G.A. Loew, S. Tantawi, J.W. Wang</i>                | 507        |
| 6.14     | HIGH VOLTAGE TECHNIQUE <i>B. Goddard</i>  | 511        |
| 6.15     | COATING RECIPES   | 515        |
| 6.15.1   | Recipes for Coating Windows <i>R.M. Sundelin, H.L. Phillips</i>   | 515        |

|          |   |            |
|----------|---|------------|
| 6.15.2   | Recipes for Coating Ceramic and Metal Vacuum Chambers <i>S.D. Henderson</i>                             | 516        |
| 6.16     | CAVITY MEASUREMENTS <i>R. Rimmer, M. Tigner</i>   | 521        |
| 6.16.1   | Field Maps by Perturbation Methods  | 521        |
| 6.16.2   | $Q$ and $\beta$ Determination from Input Coupler  | 522        |
| 6.17     | MAGNETIC MEASUREMENTS   | 523        |
| 6.17.1   | Accelerator Magnets <i>A.K. Jain, P. Wanderer</i>   | 523        |
| 6.17.2   | Insertion Device Measurement <i>S. Marks, R.D. Schlueter</i>  | 529        |
| 6.18     | HIGH POWER SWITCHES   | 533        |
| 6.18.1   | DC Switches <i>J. M. Sanders, T. Tang, M. A. Gundersen, G. Roth</i>                                     | 533        |
| 6.18.2   | Ultra-High-Power RF Switches <i>S. Tantawi</i>  | 540        |
| <b>7</b> | <b>SUBSYSTEMS</b>   | <b>543</b> |
| 7.1      | PARTICLE SOURCES  | 543        |
| 7.1.1    | Electron Sources  | 543        |
| 7.1.1.1  | Thermionic cathodes <i>H.G. Kirk</i>  | 543        |
| 7.1.1.2  | Photocathodes (unpolarized) <i>D.H. Dowell</i>  | 544        |
| 7.1.1.3  | Photocathodes (polarized) <i>T. Maruyama, M. Poelker</i>  | 546        |
| 7.1.1.4  | DC thermionic guns and preinjectors <i>H.G. Kirk</i>  | 547        |
| 7.1.1.5  | Normal conducting rf photo guns <i>D.H. Dowell</i>  | 548        |
| 7.1.1.6  | Superconducting rf photo guns <i>J. Sekutowicz</i>  | 550        |
| 7.1.1.7  | DC photo guns <i>B. Dunham</i>  | 552        |
| 7.1.1.8  | Continuous duty preinjectors <i>I. Bazarov</i>  | 553        |
| 7.1.2    | Positron Sources  | 555        |
| 7.1.2.1  | Tungsten targets <i>S. Ecklund</i>  | 555        |
| 7.1.2.2  | Conversion of undulator radiation <i>A. Mikhailichenko</i>  | 558        |
| 7.1.3    | Polarized Protons and Heavy Ions <i>T.B. Clegg, W. Haeberli</i>   | 561        |
| 7.1.4    | $H^-$ Ion Sources <i>K.N. Leung</i>   | 563        |
| 7.1.5    | Antiproton Production <i>G. Dugan</i>   | 565        |
| 7.1.6    | Multi-Charged Heavy Ion Sources <i>J. Alessi, A. Pikin</i>  | 568        |
| 7.1.7    | Foil Strippers  | 571        |
| 7.1.7.1  | Charge state strippers <i>M.A. McMahan</i>  | 571        |
| 7.1.7.2  | Stripper foils for $H^-$ beams <i>M.A. Plum</i>   | 574        |
| 7.1.8    | Lorentz Stripping of $H^-$ Ions <i>M.A. Furman</i>  | 576        |
| 7.1.9    | Laser-Assisted $H^-$ Conversion to Protons <i>V. Danilov</i>  | 576        |
| 7.2      | CONFINEMENT AND FOCUSING  | 577        |
| 7.2.1    | Resistive Magnets <i>F.E. Mills, D.J. Harding</i>   | 577        |
| 7.2.2    | Consequences of Saturation of High Permeability Material <i>K. Halbach, R. Schlueter</i>                | 584        |
| 7.2.3    | Special Topics in Magnetics   | 585        |
| 7.2.3.1  | Properties of 3D vacuum fields integrating along a straight line <i>K. Halbach, R. Schlueter</i>        | 585        |
| 7.2.3.2  | Pole width necessary to obtain desired field quality in a 2-D magnet <i>K. Halbach, R. Schlueter</i>    | 585        |
| 7.2.3.3  | Eddy currents <i>K. Halbach, R. Schlueter</i>   | 586        |
| 7.2.3.4  | Power dissipation in the dipole coils of a storage ring with iron poles <i>K. Halbach, R. Schlueter</i> | 588        |
| 7.2.3.5  | Alpha magnet <i>H. Wiedemann</i>  | 588        |
| 7.2.4    | $\cos\theta$ Superconducting Magnets <i>P. Schmüser</i>   | 588        |
| 7.2.5    | Superferric Magnets <i>A. Zeller</i>  | 595        |
| 7.2.6    | High Field Accelerator Magnets <i>G. Sabbi</i>  | 599        |
| 7.2.7    | Kickers <i>C. Burkhardt</i>   | 602        |
| 7.2.8    | Permanent Magnet Elements <i>K. Halbach, R. Schlueter</i>   | 607        |

|          |   |     |
|----------|---|-----|
| 7.2.9    | Electrostatic Separators <i>J.J. Welch</i>  | 614 |
| 7.2.10   | Deflection and Crab Cavities <i>K. Akai, M. Tigner</i>  | 617 |
| 7.2.10.1 | Introduction  | 617 |
| 7.2.10.2 | Multicell deflection cavities   | 617 |
| 7.2.10.3 | Crab cavity   | 617 |
| 7.2.11   | Electrostatic Lenses <i>A. Faltens</i>  | 620 |
| 7.2.12   | Lithium Lens <i>G. Dugan</i>  | 621 |
| 7.2.13   | Orbit Feedback Control <i>Y. Tian, L.H. Yu</i>  | 624 |
| 7.2.13.1 | Principles  | 624 |
| 7.2.13.2 | Implementation  | 626 |
| 7.2.13.3 | Local orbit feedback  | 628 |
| 7.2.14   | Feedback to Control Coupled-Bunch instabilities <i>J.D. Fox</i>   | 628 |
| 7.2.14.1 | Beam diagnostics via feedback signals   | 634 |
| 7.2.15   | Beam Deflection and Collimation with Aligned Crystals <i>R.A. Carrigan, Jr.</i>                           | 636 |
| 7.2.16   | Septum Devices <i>R. Keizer</i>   | 639 |
| 7.2.17   | Electron Lenses <i>V. Shiltsev</i>  | 641 |
| 7.2.18   | Spin Manipulation   | 641 |
| 7.2.18.1 | Siberian snake <i>V. Ptitsyn</i>  | 641 |
| 7.2.18.2 | Partial snake <i>H. Huang</i>   | 642 |
| 7.2.18.3 | Spin rotator <i>V. Ptitsyn</i>  | 643 |
| 7.2.18.4 | RF spin rotator <i>M. Bai, T. Roser</i>   | 643 |
| 7.3      | ACCELERATION  | 645 |
| 7.3.1    | RF System Design  | 645 |
| 7.3.1.1  | RF system design for stability <i>D. Boussard</i>   | 645 |
| 7.3.1.2  | Low level RF <i>J.D. Fox</i>  | 649 |
| 7.3.2    | Klystron Amplifiers   | 651 |
| 7.3.2.1  | Klystrons <i>D. Sprehn</i>  | 651 |
| 7.3.2.2  | Klystron amplifier systems <i>A. Gamp, M. Tigner</i>  | 653 |
| 7.3.3    | Tetrode Amplifiers <i>J.M. Brennan</i>  | 655 |
| 7.3.4    | Inductive Output Tube <i>J.M. Brennan, S. Belomestnykh</i>  | 659 |
| 7.3.5    | Drift Tube Linacs <i>J.M. Potter</i>  | 660 |
| 7.3.6    | Normal Conducting $v = c$ Linac Structures <i>G.A. Loew</i>   | 664 |
| 7.3.7    | Inductively Loaded Accelerating Cavities <i>M. Yoshii</i>   | 667 |
| 7.3.8    | Fixed Frequency Cavities  | 670 |
| 7.3.8.1  | Multicell cavities <i>W. Schnell</i>  | 670 |
| 7.3.8.2  | Single cell cavities <i>R. Rimmer</i>   | 672 |
| 7.3.9    | Superconducting Cavities for $v_p = c$ Linacs, Storage Rings, & Synchrotrons <i>D. Proch, H. Padamsee</i> | 674 |
| 7.3.10   | Superconducting Cavities for $v_p < c$ Linacs <i>M.P. Kelly</i>   | 681 |
| 7.3.11   | Superconducting Single Cell Cavities <i>S. Belomestnykh, J. Kirchgessner</i>                              | 683 |
| 7.3.12   | Millimeter-Wave Linacs <i>D. Whittum</i>  | 687 |
| 7.3.13   | Plasma Accelerators <i>E. Esarey, C. Schroeder</i>  | 690 |
| 7.4      | BEAM INSTRUMENTATION AND DIAGNOSTICS  | 697 |
| 7.4.1    | Composition — Ion Beams <i>P. Forck, P. Strehl</i>  | 697 |
| 7.4.2    | Longitudinal Phase Space Measurement — Ion Beams <i>P. Forck, P. Strehl</i>                               | 699 |
| 7.4.3    | Heavy Ion Linacs — Emittance Measurements <i>P. Forck, P. Strehl</i>                                      | 702 |
| 7.4.4    | Charge State — Ion Beams <i>P. Forck, P. Strehl</i>   | 704 |
| 7.4.5    | Beam Current Measurement <i>J. Hinkson</i>  | 706 |
| 7.4.6    | Beam Position Monitors <i>J.A. Hinkson, S. Smith, G. Decker</i>   | 709 |
| 7.4.7    | Longitudinal Distribution Function, Electrons <i>B. Schmidt</i>   | 713 |
| 7.4.7.1  | Longitudinal diagnostics with coherent radiation  | 713 |
| 7.4.7.2  | Electro-optical bunch length monitors   | 714 |

|          |  |            |
|----------|--|------------|
| 7.4.8    | Transverse and Longitudinal Emittance Measurements <i>J.T. Seeman</i>                                      | 716        |
| 7.4.9    | Transverse Beam Profile Measurements with Transition and Diffraction Radiation <i>P. Evtushenko</i>        | 719        |
| 7.4.10   | X-Ray Beam Size Monitor <i>J.P. Alexander, D.P. Peterson</i>   | 721        |
| 7.4.10.1 | X-ray source and optics  | 721        |
| 7.4.10.2 | Detector   | 721        |
| 7.4.10.3 | Beam size measurements   | 721        |
| 7.4.10.4 | Summary  | 723        |
| 7.4.11   | Streak Cameras <i>J.M. Byrd</i>  | 723        |
| 7.4.11.1 | Principle of operation   | 723        |
| 7.4.11.2 | Temporal resolution  | 724        |
| 7.4.11.3 | Applications to measurement of beam dynamics   | 725        |
| 7.4.12   | Laser Wire <i>M. Ross</i>  | 726        |
| 7.4.13   | Laser Interference Methods <i>T. Shintake</i>  | 728        |
| 7.4.14   | Beam Size via Synchrotron Radiation <i>T. Shintake</i>   | 730        |
| 7.4.15   | Electron Cloud Measurements <i>F. Caspers, F. Zimmermann</i>   | 732        |
| 7.4.16   | Beam Loss Monitors <i>R. Jones, K. Wittenburg</i>  | 736        |
| 7.4.16.1 | Beam loss monitoring using ionisation detection  | 737        |
| 7.4.16.2 | Beam loss monitoring using light detection   | 739        |
| 7.4.17   | Schottky Spectra   | 739        |
| 7.4.17.1 | Transverse Schottky spectra and beam transfer functions <i>O. Boine-Frankenheim, V. Kornilov, S. Paret</i> | 739        |
| 7.4.17.2 | Longitudinal Schottky spectra and beam transfer function <i>E. Shaposhnikova</i>                           | 742        |
| 7.5      | IMPEDANCE DETERMINATION  | 745        |
| 7.5.1    | Bench Measurements <i>F. Caspers</i>   | 745        |
| 7.5.2    | Beam-Based Characterization of Coupled Bunch Instabilities <i>J.M. Byrd</i>                                | 750        |
| 7.5.2.1  | Passive techniques   | 750        |
| 7.5.2.2  | Active techniques  | 751        |
| 7.5.3    | Other Beam Based Methods to Measure Impedances <i>J. Gareyte, J. S. Berg</i>                               | 752        |
| 7.5.4    | Direct Wakefield Measurement <i>W. Gai, C. Jing</i>  | 754        |
| 7.6      | POLARIMETERS   | 756        |
| 7.6.1    | Lepton Polarimeters <i>E. Chudakov, D. Gaskell, J. Grames, M. Woods</i>                                    | 756        |
| 7.6.1.1  | Mott polarimetry   | 756        |
| 7.6.1.2  | Møller polarimetry   | 756        |
| 7.6.1.3  | Compton polarimetry  | 757        |
| 7.6.2    | Proton Polarimeters <i>Y. Makdisi</i>  | 758        |
| 7.7      | CONTROLS AND TIMING <i>K. Rehlich</i>  | 760        |
| 7.8      | FEMTOSECOND PRECISION OPTICAL SYNCHRONIZATION <i>F. Löhl</i>   | 763        |
| <b>8</b> | <b>RADIATION EFFECTS AND PROTECTION</b>  | <b>767</b> |
| 8.1      | RADIATION PROTECTION PRINCIPLES <i>S. Roesler, M. Silari</i>   | 767        |
| 8.1.1    | Quantities   | 767        |
| 8.1.1.1  | Physical quantities  | 767        |
| 8.1.1.2  | Protection quantities  | 767        |
| 8.1.1.3  | Operational quantities   | 768        |
| 8.1.1.4  | Dose conversion coefficients   | 768        |
| 8.1.2    | Health Effects of Ionizing Radiation   | 768        |
| 8.1.3    | Radiation Levels   | 769        |
| 8.2      | SOURCES OF RADIATION <i>R.H. Thomas, S. Roesler, M. Silari</i>   | 769        |
| 8.2.1    | Electron Beams   | 770        |
| 8.2.1.1  | Bremsstrahlung   | 770        |

|         |  |     |
|---------|--|-----|
| 8.2.1.2 | Electromagnetic cascade . . . . .  | 771 |
| 8.2.2   | Proton Beams . . . . .   | 773 |
| 8.2.3   | Ion Beams . . . . .  | 774 |
| 8.3     | MONTE CARLO CODES <i>S. Roesler, M. Silari</i> . . . . .                 | 776 |
| 8.3.1   | FLUKA . . . . .  | 776 |
| 8.3.2   | GEANT4 . . . . .   | 776 |
| 8.3.3   | MARS15 . . . . .   | 777 |
| 8.3.4   | MCNPX . . . . .  | 778 |
| 8.3.5   | PHITS . . . . .  | 778 |
| 8.4     | RADIATION TRANSPORT AND SHIELDING <i>S. Roesler, M. Silari</i> . . . . . | 779 |
| 8.4.1   | General Concepts . . . . .   | 779 |
| 8.4.2   | Semi-Empirical Models . . . . .  | 779 |
| 8.4.3   | The Monte Carlo “hybrid” Method . . . . .                                | 780 |
| 8.4.4   | Photon Transmission . . . . .  | 781 |
| 8.4.5   | Ducts and Labyrinths for Neutron Attenuation . . . . .                   | 782 |
| 8.4.6   | Ducts and Labyrinths for Photon Attenuation . . . . .                    | 783 |
| 8.4.7   | Shielding Materials . . . . .  | 784 |
| 8.4.8   | Muons . . . . .  | 785 |
| 8.4.9   | Skyshine . . . . .   | 785 |
| 8.5     | RADIATION MEASUREMENTS <i>S. Roesler, M. Silari</i> . . . . .            | 786 |
| 8.5.1   | Neutron Dosimetry and Spectrometry . . . . .                             | 786 |
| 8.5.1.1 | Rem counters . . . . .   | 786 |
| 8.5.1.2 | Bonner sphere spectrometers . . . . .                                    | 787 |
| 8.5.1.3 | Bubble detectors . . . . .   | 787 |
| 8.5.1.4 | Track etched detectors . . . . .   | 788 |
| 8.5.2   | Photon Dosimetry and Spectrometry . . . . .                              | 788 |
| 8.5.2.1 | BC501A liquid scintillator . . . . .                                     | 789 |
| 8.5.2.2 | BGO photon spectrometer . . . . .  | 789 |
| 8.5.2.3 | Geiger Müller counter . . . . .  | 789 |
| 8.5.3   | Detectors Based on Micro-Dosimetric Principles . . . . .                 | 789 |
| 8.6     | INDUCED RADIOACTIVITY <i>S. Roesler, M. Silari</i> . . . . .             | 790 |
| 8.6.1   | Fundamental Principles . . . . .   | 790 |
| 8.6.2   | Solid Materials . . . . .  | 791 |
| 8.6.3   | Liquids . . . . .  | 792 |
| 8.6.4   | Air . . . . .  | 793 |
| 8.7     | RADIATION DAMAGE THRESHOLDS <i>H. Schönbacher</i> . . . . .              | 793 |
| 8.7.1   | Organic Materials . . . . .  | 793 |
| 8.7.2   | Semiconductors and Electronic Devices . . . . .                          | 797 |
| 8.7.3   | Ceramics . . . . .   | 799 |
| 8.7.4   | Vitreous Materials . . . . .   | 799 |
| 8.7.5   | Metals . . . . .   | 801 |
| 8.7.6   | Summary . . . . .  | 801 |

## Author index

I-1

## Subject Index

I-5

# Chapter 1. INTRODUCTION

## 1.1 HOW TO USE THIS BOOK

This is an accelerator designer's and operator's handbook of formulae, tables, figures and references. It is meant to be a concise working tool. An effort has been made to provide an index which is as complete as possible.

Each subsection (e.g. 2.3.4) is treated as a unit which is more or less self-contained. Numbering of all figures and tables are reset at each subsection, and references are found following each subsection. References are not meant to be exhaustive but represent the experts' recommendation about a reliable place to begin.

While the linear and circular accelerators for high energy physics and synchrotron radiation applications are our primary concern, we have tried to provide connections to other types of accelerators in the glossary section 1.6.

## 1.2 NOMENCLATURE

Boldface symbol means matrix quantity.  $\mathbf{A}^t$  is the transpose of  $\mathbf{A}$ . A tilde,  $\tilde{A}$ , means Fourier transform.  $\vec{A}$  is a vector.  $\hat{a}$  is a unit vector.

For complex numbers, we use  $i = \sqrt{-1}$ , and a sinusoidal time dependence is described by  $e^{-i\omega t}$  unless otherwise noted.

### Abbreviations

|                |                                     |
|----------------|-------------------------------------|
| ac, dc         | alternating current, direct current |
| BBU            | beam break-up (instability)         |
| BPM            | beam position monitor               |
| CM             | center of mass                      |
| c.w.           | continuous wave                     |
| EM             | electromagnetic                     |
| FEL            | free electron laser                 |
| FWHM           | full-width at half-maximum          |
| HOM            | higher order mode                   |
| IR, UV         | infrared, ultraviolet               |
| IP, IR         | interaction point, region           |
| LHe            | liquid helium                       |
| l.h.s., r.h.s. | left-hand side, right-hand side     |
| $n$ -D         | $n$ -dimension(al)                  |
| rf             | radio frequency                     |

|        |                               |
|--------|-------------------------------|
| NC     | normal conducting             |
| RFQ    | rf quadrupole                 |
| rms    | root mean square              |
| RRR    | residual resistivity ratio    |
| SC     | superconducting               |
| SR     | synchrotron radiation         |
| SRF    | superconducting rf            |
| SW, TW | standing wave, traveling wave |
| UHV    | ultrahigh vacuum              |

### Abbreviations of references

|         |   |
|---------|---|
| APAC    | Asian PAC   |
| APL     | Applied Physics Letters                               |
| EPAC    | European PAC  |
| IJMP    | International Journal of Modern Physics               |
| IPAC    | International PAC                                     |
| JAP     | Journal of Applied Physics                            |
| JETP    | Journal of Experimental and Theoretical Physics       |
| JINST   | Journal of Instrumentation                            |
| JVST    | Journal of Vacuum Science and Technology              |
| NIM     | Nuclear Instruments & Methods                         |
| NJP     | New Journal of Physics                                |
| PA      | Particle Accelerators                                 |
| PAC     | Particle Accelerator Conference                       |
| PL      | Physics Letters                                       |
| PR      | Physical Review                                       |
| PRL     | Physical Review Letters                               |
| PRST-AB | Physical Review Special Topics – Accelerators & Beams |
| RAST    | Reviews of Accelerator Science & Technology           |
| RMP     | Reviews of Modern Physics                             |
| RSI     | Review of Scientific Instruments                      |

Tabulated below are symbols adopted throughout this Handbook unless otherwise noted. More symbols are defined in the table of fundamental constants in Sec.1.3. and locally in the text.

| Symbol               | Quantity   |
|----------------------|--|
| $\alpha_p$           | momentum compaction factor   |
| $\alpha_T$           | thermal expansion coefficient  |
| $\alpha_{x,y}$       | horizontal ( $x$ ), vertical ( $y$ ) Courant-Snyder alpha functions                        |
| $\mathcal{B}$        | brightness   |
| $\vec{B}$            | magnetic induction   |
| $(B\rho)$            | $= P/e$ , rigidity of a particle of charge $e$<br>$(B\rho)$ [T-m] = 3.335641 $P_0$ [GeV/c] |
| $\beta, \vec{\beta}$ | $= (v, \vec{v})/c$ , (speed, velocity) relative to light                                   |

## Sec.1.2: NOMENCLATURE

|                     |  |  |   |
|---------------------|--|--|---|
| $\beta_{x,y}$       | horizontal ( $x$ ), vertical ( $y$ )   | $\Omega$   | solid angle   |
| $C$                 | Courant-Snyder beta functions  | $P, \vec{P}, p, \vec{p}$   | particle momentum   |
| $c_v, c_p$          | circumference  | $P_0, p_0$   | design particle momentum  |
| $\bar{D}$           | specific heat  | $P$  | power   |
| $\delta$            | = $\Delta P/P_0$ , relative momentum error   | $Q$  | quality factor of oscillator  |
| $D_{x,y}$           | horizontal ( $x$ ), vertical ( $y$ )   | $q$  | charge on a particle  |
|                     | dispersion functions   | $R$  | = $C/(2\pi)$ , average radius   |
| $\mathcal{D}_{x,y}$ | beam-beam disruption parameters  | $\rho$   | bending radius  |
| $\det \mathbf{A}$   | determinant of matrix $\mathbf{A}$   | $\rho$   | volume density  |
| $E$                 | Young's modulus  | $\rho_r$   | resistivity   |
| $E$                 | particle total energy  | $s$  | longitudinal coordinate   |
| $E_0$               | design particle energy   | $\sigma$   | along an accelerator  |
| $E_0$               | = $mc^2$ rest mass energy  | $\sigma_c$   | interaction cross-section   |
| $\vec{E}$           | electric field   | $\sigma_{x,y,z}$   | = $1/\rho_r$ , conductivity   |
| $\epsilon_{x,y,L}$  | horizontal ( $x$ ), vertical ( $y$ ),<br>longitudinal ( $L$ ) emittances<br>(unnormalized) | $\sigma_{x',y'}$   | horiz. ( $x$ ), vert. ( $y$ ),<br>long. ( $z$ ) rms beam size                   |
| $\epsilon_{Nx,Ny}$  | = $\beta\gamma\epsilon_{x,y}$ , normalized emittances                                      | $\sigma_{E,\delta}$  | horiz. ( $x'$ ), vert. ( $y'$ )   |
| $\eta$              | = $\alpha_p - (1/\gamma^2)$ , phase slip factor  | $\text{sgn}(x)$  | rms angular spread  |
| $f_0$               | revolution frequency   | $t$  | energy ( $E$ ), relative  |
| $\gamma$            | = $1/\sqrt{1 - \beta^2}$ , relativistic factor   | $T$  | momentum ( $\delta$ ) rms spread  |
| $\gamma_{x,y}$      | horizontal ( $x$ ), vertical ( $y$ )   | $T$  | sign function,  |
|                     | Courant-Snyder gamma functions   |  | = 1 if $x > 0$ , = -1 if $x < 0$  |
| $\gamma_t$          | transition gamma   | $t$  | time  |
| $H$                 | Hamiltonian  | $T$  | temperature   |
| $\mathcal{H}$       | synchrotron integral $H$ -function   | $\text{tr} \mathbf{A}$   | kinetic energy  |
| $h$                 | rf harmonic number   | $T_0$  | trace of matrix $\mathbf{A}$  |
| $\vec{H}$           | magnetic field   | $U_0$  | revolution period   |
| $H(x)$              | step function,<br>= 1 if $x > 0$ , = 0 if $x < 0$  | $\Upsilon$   | synchrotron radiation loss  |
| $I_0$               | = $N_B e f_0$ , average bunch current  | $v, \vec{v}$   | per revolution  |
| $I_b$               | peak bunch current   | $v_g$  | beamstrahlung parameter   |
| $J_{x,y,s}$         | horizontal ( $x$ ), vertical ( $y$ ),<br>synchrotron ( $s$ ) partition numbers             | $v_p$  | speed, velocity   |
| $\vec{J}$           | current density  | $V_{\text{rf}}$  | group velocity  |
| $k$                 | = $2\pi/\lambda$ , wave number   | $W_{\parallel m}, W_{\perp m}$<br>(or $W_m^{\parallel}, W_m^{\perp}$ ) | phase velocity  |
| $\kappa$            | thermal conductivity   | $x$  | rf voltage  |
| $\mathcal{L}$       | luminosity   | $x'$   | longitudinal, transverse  |
| $\lambda$           | wavelength   | $X_0$  | wake function of mode $m$   |
| $\mu$               | Poisson's ratio  | $\xi_{x,y}$  | ( $W_{\parallel}$ if $m = 0$ , $W_{\perp}$ if $m = 1$ )                         |
| $\mu$               | betatron phase advance per<br>turn or period   | $y$  | horiz. displacement   |
| $N$                 | total no. of particles in beam   | $y'$   | = $dx/ds$ , horiz. angular  |
| $N_B$               | number of particles per bunch<br>in a bunched beam   | $\psi$   | deviation of a particle   |
| $n_B$               | number of bunches in beam  | $\psi_{x,y}$   | radiation length  |
| $\nu_{x,y,s}$       | horizontal betatron ( $x$ ),<br>vertical betatron ( $y$ ),<br>synchrotron ( $s$ ) tune     | $z$  | = $(d\nu_{x,y}/d\delta)$ , horizontal ( $x$ ),<br>vertical ( $y$ ) chromaticity |
| $\nu_{\text{sp}}$   | spin tune  |  | vert. displacement  |
| $\omega$            | = $2\pi f$ , angular frequency   |  | = $dy/ds$ , vert. angular   |
| $\omega_0$          | = $2\pi f_0$ , angular revolution freq.  |  | deviation of a particle   |
| $\omega_{x,y,s}$    | = $\nu_{x,y,s}\omega_0$ , betatron ( $x, y$ ),<br>synchrotron ( $s$ ) angular freq.        |  | distribution density in phase   |
|                     |  |  | space, normalized to unity  |
|                     |  |  | horiz. ( $x$ ), vert. ( $y$ )   |
|                     |  |  | betatron phase  |
|                     |  |  | long. displacement of a particle  |
|                     |  |  | relative to synchronous particle  |
|                     |  |  | ( $z > 0$ ahead, $z < 0$ behind)  |
|                     |  | $Z_{\parallel m}, Z_{\perp m}$<br>(or $Z_m^{\parallel}, Z_m^{\perp}$ ) | longitudinal, transverse  |
|                     |  |  | impedance of mode $m$   |
|                     |  |  | ( $Z_{\parallel}$ if $m = 0$ , $Z_{\perp}$ if $m = 1$ )                         |

### 1.3 FUNDAMENTAL CONSTANTS [1]

| Quantity                     | Symbol  | Value   |
|------------------------------|---|---|
| pi                           | $\pi$   | 3.141592653589793238                                |
| exponential constant         | $e$   | 2.718281828459045235                                |
| Euler's constant             | $\gamma$                                      | 0.5772156649  |
| speed of light               | $c$   | 2.99792458 E8 m s <sup>-1</sup> (exact)             |
| permeability of vacuum       | $\mu_0$                                       | 4 $\pi$ E-7 Henry m <sup>-1</sup> (exact)           |
| permittivity of vacuum       | $\epsilon_0 = 1/(\mu_0 c^2)$                  | 8.854187817 E-12 Farad m <sup>-1</sup>              |
| electronic charge            | $e$   | 1.6021765 E-19 C = 4.8032043 E-10 esu               |
| Planck constant              | $h$   | 6.626069 E-34 J s                                   |
| reduced Planck constant      | $\hbar = h/(2\pi)$                            | 1.054572 E-34 J s = 6.582119 E-16 eV s              |
| Boltzmann constant           | $k_B$   | 1.380650 E-23 J K <sup>-1</sup>                     |
| Avogadro number              | $N_A$   | 6.022142 E23 mole <sup>-1</sup>                     |
| gravitational constant       | $G$   | 6.67428 E-11 Newton m <sup>2</sup> kg <sup>-2</sup> |
| std. grav. accel.            | $g$   | 9.80665 m s <sup>-2</sup>                           |
| electron mass                | $m_e$   | 9.1093822 E-31 kg                                   |
| proton mass                  | $m_p$   | 1.6726216 E-27 kg                                   |
| rest mass energy of          |   |   |
| electron                     | $m_e c^2$                                     | 0.51099891 MeV                                      |
| proton                       | $m_p c^2$                                     | 938.27201 MeV                                       |
| neutron                      | $m_n c^2$                                     | 939.5653 MeV  |
| deuteron                     | $m_d c^2$                                     | 1875.6128 MeV                                       |
| muon                         | $m_\mu c^2$                                   | 105.65837 MeV                                       |
| Z-particle                   | $m_Z c^2$                                     | 91.188 GeV  |
| W-particle                   | $m_W c^2$                                     | 80.399 GeV  |
| anomalous gyromagnetic ratio | $G = (g - 2)/2$                               | 0.00115965219                                       |
| electron                     |   | 0.001165923   |
| muon                         |   | 1.79284739  |
| proton                       |   | -0.1429878  |
| deuteron                     |   |   |
| fine structure constant      | $\alpha_F = e^2/(4\pi\epsilon_0\hbar c)$      | 1/137.0359997                                       |
| impedance of free space      | $Z_0 = \sqrt{\mu_0/\epsilon_0} = \mu_0 c$     | 376.7303 Ω  |
| classical radius of          |   |   |
| electron                     | $r_e = e^2/(4\pi\epsilon_0 m_e c^2)$          | 2.8179403 E-15 m                                    |
| proton                       | $r_p = e^2/(4\pi\epsilon_0 m_p c^2)$          | 1.534698 E-18 m                                     |
| electron Compton wavelength  | $\lambda_e = h/(m_e c)$                       | 2.4263106 E-12 m                                    |
|                              | $\lambda_e = \lambda_e/(2\pi)$                | 0.3861593 E-12 m                                    |
| Alfvén current               | $I_A = ec/r_e$                                | 17.045093 kA  |
| Bohr radius                  | $a_\infty = 4\pi\epsilon_0\hbar^2/(m_e e^2)$  | 5.29177209 E-11 m                                   |
| Thomson cross section        | $\sigma_T = (8\pi/3)r_e^2$                    | 6.65246 E-29 m <sup>2</sup>                         |
| Bohr magneton                | $\mu_B = e\hbar/(2m_e c)$                     | 5.7883818 E-5 eV/Tesla                              |
| nuclear magneton             | $\mu_N = e\hbar/(2m_p c)$                     | 3.1524512 E-8 eV/Tesla                              |
| Stefan-Boltzmann constant    | $\sigma_{SB} = (\pi^2/60)k_B^4/(\hbar^3 c^2)$ | 5.67040 E-8 W m <sup>-2</sup> K <sup>-4</sup>       |
| gas constant                 | $R = N_A k_B$                                 | 8.3145 J K <sup>-1</sup> mole <sup>-1</sup>         |

### References

- [1] K. Nakamura et al. (Particle Data Group), J. Phys. G37, 075021 (2010)

## 1.4 UNITS AND CONVERSIONS

### 1.4.1 Units

*A.W. Chao, SLAC*

We use the SI (Système International, MKSA) units throughout this Handbook unless otherwise noted. Table below gives the conversion of various physical quantities from Gaussian to SI unit systems [1].

| Quantity            | Gaussian   | SI                                       |
|---------------------|------------|--|
| speed of light      | $c$        | $\frac{1}{\sqrt{\mu_0 \epsilon_0}}$      |
| charge              | $q$        | $\frac{q}{\sqrt{4\pi\epsilon_0}}$        |
| charge density      | $\rho$     | $\frac{\rho}{\sqrt{4\pi\epsilon_0}}$     |
| current             | $I$        | $\frac{I}{\sqrt{4\pi\epsilon_0}}$        |
| current density     | $\vec{J}$  | $\frac{\vec{J}}{\sqrt{4\pi\epsilon_0}}$  |
| scalar potential    | $\Phi$     | $\sqrt{4\pi\epsilon_0}\Phi$              |
| vector potential    | $\vec{A}$  | $\sqrt{\frac{4\pi}{\mu_0}}\vec{A}$       |
| voltage             | $V$        | $\sqrt{4\pi\epsilon_0}V$                 |
| electric field      | $\vec{E}$  | $\sqrt{4\pi\epsilon_0}\vec{E}$           |
| displacement        | $\vec{D}$  | $\sqrt{\frac{\epsilon_0}{\mu_0}}\vec{D}$ |
| magnetic induction  | $\vec{B}$  | $\sqrt{\frac{4\pi}{\mu_0}}\vec{B}$       |
| magnetic field      | $\vec{H}$  | $\sqrt{4\pi\mu_0}\vec{H}$                |
| conductivity        | $\sigma_c$ | $\frac{\sigma_c}{4\pi\epsilon_0}$        |
| dielectric constant | $\epsilon$ | $\epsilon/\epsilon_0$                    |
| permeability        | $\mu$      | $\frac{\mu}{\mu_0}$                      |
| resistance          | $R$        | $4\pi\epsilon_0 R$                       |
| inductance          | $L$        | $4\pi\epsilon_0 L$                       |
| capacitance         | $C$        | $\frac{C}{4\pi\epsilon_0}$               |

Table below gives some numerical conversions between Gaussian and SI units.

| Quantity     | Gaussian                        | SI                 |
|--------------|---------------------------------|--------------------|
| Conductivity | $8.9876\text{E}9\text{ s}^{-1}$ | $= 1\text{ mho/m}$ |
| Resistance   | $1.1127\text{E}-12\text{ s/cm}$ | $= 1\text{ ohm}$   |
| Capacitance  | $8.9876\text{E}11\text{ cm}$    | $= 1\text{ farad}$ |
| Inductance   | $1.1127\text{E}-12$             | $= 1\text{ henry}$ |

If  $P$  is power and  $V$  is voltage, then

$$\text{dB} = 10 \log_{10}(P_1/P_2), \text{ or } 20 \log_{10}(V_1/V_2)$$

Other notations: dBm (reference point 1 milliwatt); dBc (reference to the carrier power); dBA is used in acoustics in reference to a standard sound pressure of  $2\text{E}-4$  microbar.

### References

- [1] J.D. Jackson, Classical Electrodynamics, 3rd ed., Wiley (1999)

### 1.4.2 Conversions

*M. Tigner, Cornell U.*

key:

**Quantity**, symbol, name in SI, Abbr., [dim.]

Quantity to be converted, conversion factor

Quantity in SI = Quantity to be converted  $\times$  conversion factor, E1 = 10, E2 = 100, etc.

**Length**,  $\ell$ , meter, m, [L]

|                |          |
|----------------|----------|
| inch, in.      | 2.54E-2  |
| foot, ft.      | 0.3048   |
| angstrom, Å    | 1.0E-10  |
| fermi, fm      | 1.0E-15  |
| light year, ly | 0.946E16 |

**Area**,  $A$ , sq. meter, m<sup>2</sup>, [L<sup>2</sup>]

|                           |           |
|---------------------------|-----------|
| sq. inch, in <sup>2</sup> | 6.4516E-4 |
| sq. foot, ft <sup>2</sup> | 9.2903E-2 |
| acre                      | 4.0468E3  |
| hectare, ha               | 1E4       |
| sq. mile, mi <sup>2</sup> | 2.5888E6  |
| barn, bn                  | 1E-28     |

**Volume**,  $V$ , cu. meter, m<sup>3</sup>, [L<sup>3</sup>]

|                           |          |
|---------------------------|----------|
| cu. inch, in <sup>3</sup> | 1.639E-5 |
| gallon, gal. (liquid)     | 3.785E-3 |
| cu. ft., ft <sup>3</sup>  | 2.832E-2 |

**Mass**,  $M$ , kilogram, kg, [M]

|      |       |
|------|-------|
| slug | 14.59 |
|------|-------|

**Density**,  $\rho$ , kg/m<sup>3</sup>, [M/L<sup>3</sup>]

|                                 |         |
|---------------------------------|---------|
| slug/cu ft                      | 515.4   |
| pound/cu in, lb/in <sup>3</sup> | 2.768E4 |

**Time**,  $t$ , second, s, [T]

|           |         |
|-----------|---------|
| year, yr. | 3.156E7 |
|-----------|---------|

**Speed**,  $v$ , meter/sec, m/s, [L/T]

|                  |        |
|------------------|--------|
| foot/sec, ft/s   | 0.3048 |
| mile/hour, mi/hr | 0.4470 |

**Force**,  $F$ , Newton, N, [ML/T<sup>2</sup>]

|           |        |
|-----------|--------|
| dyne      | 1.0E-5 |
| pound, lb | 4.448  |

**Pressure**,  $P$ , Pascal, Pa, N/m<sup>2</sup>, [M/T<sup>2</sup>L]

|                 |         |
|-----------------|---------|
| atmosphere, atm | 1.013E5 |
|-----------------|---------|

|     |       |
|-----|-------|
| bar | 1.0E5 |
|-----|-------|

|                      |     |
|----------------------|-----|
| dyne/cm <sup>2</sup> | 0.1 |
|----------------------|-----|

|                             |         |
|-----------------------------|---------|
| pound/in <sup>2</sup> , psi | 6.895E3 |
|-----------------------------|---------|

|                            |         |
|----------------------------|---------|
| in. H <sub>2</sub> O @ 4°C | 2.491E2 |
|----------------------------|---------|

|                  |          |
|------------------|----------|
| Torr, mmHg @ 0°C | 1.3332E2 |
|------------------|----------|

**Energy**,  $W$ , Joule, J, [ML<sup>2</sup>/T<sup>2</sup>]

|     |      |
|-----|------|
| BTU | 1055 |
|-----|------|

|     |        |
|-----|--------|
| erg | 1.0E-7 |
|-----|--------|

|                   |       |
|-------------------|-------|
| foot-pound, ft lb | 1.356 |
|-------------------|-------|

|                 |         |
|-----------------|---------|
| horsepower hour | 2.685E6 |
|-----------------|---------|

|   |           |
|---|-----------|
| calorie, cal  | 4.186     |
| kilowatt hour, kWhr   | 3.6E6     |
| electron volt, eV   | 1.602E-19 |
| liter atmosphere  | 101.31    |
| <b>Power, <math>P</math>, watt, W, [ML<sup>2</sup>/T<sup>3</sup>]</b>                           |           |
| erg/s   | 1.0E-5    |
| BTU/hr  | 0.2930    |
| foot pound/sec, ft lb/s   | 1.356     |
| horsepower  | 745.7     |
| calorie/sec, cal/s  | 4.186     |
| <b>Thermal conductivity, <math>\kappa</math>, W/m K, [ML/T<sup>3</sup>]</b>                     |           |
| Watt/cm K   | 1.0E2     |
| BTU/hr/ft <sup>2</sup> /°F/ft   | 1.73      |
| cal/cm·s  | 418.6     |
| <b>Specific heat, <math>c_{v,p}</math>, J/kg K, [L<sup>2</sup>/T<sup>2</sup>]</b>               |           |
| cal/gm °C   | 4.186E3   |
| BTU/lb °F   | 4.186E3   |
| <b>Viscosity, <math>\eta</math>, kg/m·s, [M/LT]</b>   |           |
| poises, g/cm·s  | 0.1       |
| slug/ft·s   | 4.79E1    |
| <b>Charge, <math>q</math>, Coulomb, C, [M<sup>1/2</sup>L<sup>1/2</sup>]</b>                     |           |
| abcoulomb   | 10        |
| statcoulomb   | 3.336E-10 |
| <b>Current, <math>I</math>, Ampere, A, [M<sup>1/2</sup>L<sup>1/2</sup>/T]</b>                   |           |
| abampere  | 10        |
| statampere  | 3.336E-10 |
| <b>Potential, <math>V</math>, volt, V, [M<sup>1/2</sup>L<sup>3/2</sup>/T<sup>2</sup>]</b>       |           |
| abvolt  | 1.0E-8    |
| statvolt  | 2.9979E2  |
| <b>Elec. Field, <math>E</math>, volt/m, V/m, [M<sup>1/2</sup>L<sup>1/2</sup>/T<sup>2</sup>]</b> |           |
| abvolt/cm   | 1.0E-6    |
| statvolt/cm   | 2.9979E4  |
| <b>Mag. Field, <math>H</math>, A-turn/m, [M<sup>1/2</sup>/L<sup>1/2</sup>T]</b>                 |           |
| Oersted   | 79.58     |
| <b>Displacement, <math>D</math>, Coul/m<sup>2</sup>, [M<sup>1/2</sup>/L<sup>3/2</sup>]</b>      |           |
| abcoulomb/cm <sup>2</sup>   | 7.958E3   |
| statcoulomb/cm <sup>2</sup>   | 2.654E-7  |
| <b>Mag. flux, <math>\Phi</math>, Weber, volt-s, [M<sup>1/2</sup>L<sup>3/2</sup>/T]</b>          |           |
| Tesla m <sup>2</sup>  | 1         |
| maxwell, abvolt-s   | 1.0E-8    |
| statweber, statvolt-s   | 2.9979    |
| <b>Flux density, <math>B</math>, Tesla,T, [M<sup>1/2</sup>/L<sup>1/2</sup>T]</b>                |           |
| Gauss   | 1.0E-4    |
| esu   | 2.9979E6  |
| <b>Conductivity, <math>\sigma_c</math>, siemens/m, S/m [T/L<sup>2</sup>]</b>                    |           |
| mho/cm  | 1.0E2     |
| <b>Resistivity, <math>\rho_r</math>, ohm meter, Ω-m, [L<sup>2</sup>/T]</b>                      |           |
| microhm cm  | 1.0E-8    |
| <b>Resistance, <math>R</math>, ohm, Ω, [L/T]</b>  |           |
| abohm   | 1.0E-9    |
| stathohm  | 8.987E11  |

|   |           |
|---|-----------|
| <b>Capacitance, <math>C</math>, farad, f, [T<sup>2</sup>/L]</b> |           |
| abfarad   | 1.0E9     |
| statfarad   | 1.112E-11 |

|   |           |
|---|-----------|
| <b>Inductance, <math>L</math>, henry, hy, [L]</b> |           |
| abhenry   | 1.0E-9    |
| sthathenry  | 8.9874E11 |

|                                      |        |
|--------------------------------------|--------|
| <b>Activity, bequerel, Bq, [1/T]</b> |        |
| Curie, Ci                            | 3.7E10 |

|  |        |
|--|--------|
| <b>Dose, gray, Gy, [L<sup>2</sup>/T<sup>2</sup>]</b> |        |
| rad  | 1.0E-2 |

**Exposure**, see Ch.8

**Temperature,  $T$ , Kelvin, K**

$$T[\text{°C}] = T[\text{K}] - 273.16$$

$$T[\text{°F}] = 9/5(T[\text{°C}]) + 32$$

## 1.5 FUNDAMENTAL FORMULAE

A.W. Chao, SLAC

### 1.5.1 Special Functions

Error function  $\text{erf}(x)$ ,  $\text{erfc}(x) = 1 - \text{erf}(x)$ :

$$\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x dt e^{-t^2}, \quad \text{erf}(\infty) = 1$$

Bessel functions  $J_\nu(x)$ ,  $I_\nu(x)$ ,  $N_\nu(x)$ ,  $K_\nu(x)$ :

$$\begin{aligned} & \frac{d^2 J_\nu}{dx^2} + \frac{1}{x} \frac{dJ_\nu}{dx} + \left(1 - \frac{\nu^2}{x^2}\right) J_\nu = 0 \\ & e^{-ix \cos \theta} = \sum_{n=-\infty}^{\infty} i^{-n} J_n(x) e^{inx} \\ & J_n(x) = \frac{1}{2\pi} \int_0^{2\pi} d\theta e^{-in\theta + ix \sin \theta} \\ & = \frac{i^n}{2\pi} \int_0^{2\pi} d\theta e^{in\theta - ix \cos \theta} \\ & I_n(x) = i^{-n} J_n(ix) \end{aligned}$$

$$J_n(-x) = J_{-n}(x) = (-1)^n J_n(x)$$

$$I_n(-x) = (-1)^n I_{-n}(x) = (-1)^n I_n(x)$$

$$J_n(x) \approx I_n(x) \approx \frac{1}{n!} (x/2)^n \quad \text{for } |x| \ll 1$$

$$N_\nu(x) = \frac{J_\nu(x) \cos \nu\pi - J_{-\nu}(x)}{\sin \nu\pi}$$

$$K_\nu(x) = \frac{\pi}{2 \sin \nu\pi} [I_{-\nu}(x) - I_\nu(x)]$$

$$J'_0 = -J_1, \quad N'_0 = -N_1$$

$$I'_0 = I_1, \quad K'_0 = -K_1$$

For  $x \gg 1$ ,

$$J_\nu(x) \approx \sqrt{\frac{2}{\pi x}} \cos\left(x - \frac{\pi}{2}\nu - \frac{\pi}{4}\right)$$

$$N_\nu(x) \approx \sqrt{\frac{2}{\pi x}} \sin\left(x - \frac{\pi}{2}\nu - \frac{\pi}{4}\right)$$

$$I_\nu(x) \approx \frac{1}{\sqrt{2\pi x}} e^x, \quad K_\nu(x) \approx \frac{1}{\sqrt{2\pi x}} e^{-x}$$

For roots of  $J_n$  and  $J'_n$ , see Sec.6.6.

Gamma function  $\Gamma(x)$ :

$$\Gamma(x > 0) = \int_0^\infty dt e^{-tx - t^2}$$

$$\Gamma(n+1) = n!, \quad \Gamma(x+1) = x\Gamma(x)$$

$$\Gamma(x)\Gamma(1-x) = \frac{\pi}{\sin \pi x}$$

$$\Gamma(1/2) = \sqrt{\pi}, \quad \Gamma(1/4) \approx 3.626$$

### 1.5.2 Curvilinear Coordinate Systems

General Orthogonal System ( $u_1, u_2, u_3$ ) [1]:

| System        | $u_1$ | $u_2$    | $u_3$  | $h_1$ | $h_2$ | $h_3$                |
|---------------|-------|----------|--------|-------|-------|----------------------|
| Cartesian     | $x$   | $y$      | $z$    | 1     | 1     | 1                    |
| Cylindrical   | $r$   | $\theta$ | $z$    | 1     | $r$   | 1                    |
| Spherical     | $r$   | $\theta$ | $\phi$ | 1     | $r$   | $r \sin \theta$      |
| Frenet-Serret | $x$   | $y$      | $s$    | 1     | 1     | $1 + \frac{x}{\rho}$ |

$$\begin{aligned} d\vec{s} &\equiv h_1 du_1 \hat{u}_1 + h_2 du_2 \hat{u}_2 + h_3 du_3 \hat{u}_3 \\ ds^2 &= h_1^2 du_1^2 + h_2^2 du_2^2 + h_3^2 du_3^2 \\ dV &= h_1 h_2 h_3 du_1 du_2 du_3 \\ \nabla \psi &= \frac{1}{h_1} \frac{\partial \psi}{\partial u_1} \hat{u}_1 + \frac{1}{h_2} \frac{\partial \psi}{\partial u_2} \hat{u}_2 + \frac{1}{h_3} \frac{\partial \psi}{\partial u_3} \hat{u}_3 \\ \nabla \cdot \vec{A} &= \frac{1}{h_1 h_2 h_3} \left[ \frac{\partial}{\partial u_1} (h_2 h_3 A_1) + \frac{\partial}{\partial u_2} (h_3 h_1 A_2) + \frac{\partial}{\partial u_3} (h_1 h_2 A_3) \right] \\ \nabla \times \vec{A} &= \frac{1}{h_1 h_2 h_3} \times \\ &\left\{ h_1 \hat{u}_1 \left[ \frac{\partial}{\partial u_2} (h_3 A_3) - \frac{\partial}{\partial u_3} (h_2 A_2) \right] + h_2 \hat{u}_2 \left[ \frac{\partial}{\partial u_3} (h_1 A_1) - \frac{\partial}{\partial u_1} (h_3 A_3) \right] + h_3 \hat{u}_3 \left[ \frac{\partial}{\partial u_1} (h_2 A_2) - \frac{\partial}{\partial u_2} (h_1 A_1) \right] \right\} \\ \nabla^2 \psi &= \frac{1}{h_1 h_2 h_3} \left[ \frac{\partial}{\partial u_1} \left( \frac{h_2 h_3}{h_1} \frac{\partial \psi}{\partial u_1} \right) + \frac{\partial}{\partial u_2} \left( \frac{h_3 h_1}{h_2} \frac{\partial \psi}{\partial u_2} \right) + \frac{\partial}{\partial u_3} \left( \frac{h_1 h_2}{h_3} \frac{\partial \psi}{\partial u_3} \right) \right] \end{aligned}$$

Cartesian:

$$\begin{aligned} ds^2 &= dx^2 + dy^2 + dz^2 \\ dV &= dx dy dz \\ \nabla \psi &= \frac{\partial \psi}{\partial x} \hat{x} + \frac{\partial \psi}{\partial y} \hat{y} + \frac{\partial \psi}{\partial z} \hat{z} \\ \nabla \cdot \vec{A} &= \frac{\partial A_x}{\partial x} + \frac{\partial A_y}{\partial y} + \frac{\partial A_z}{\partial z} \\ \nabla \times \vec{A} &= \hat{x} \left( \frac{\partial A_z}{\partial y} - \frac{\partial A_y}{\partial z} \right) + \hat{y} \left( \frac{\partial A_x}{\partial z} - \frac{\partial A_z}{\partial x} \right) + \hat{z} \left( \frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y} \right) \\ \nabla^2 \psi &= \frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} + \frac{\partial^2 \psi}{\partial z^2} \end{aligned}$$

Cylindrical:

$$\begin{aligned} ds^2 &= dr^2 + r^2 d\theta^2 + dz^2 \\ dV &= r dr d\theta dz \\ \nabla \psi &= \frac{\partial \psi}{\partial r} \hat{r} + \frac{1}{r} \frac{\partial \psi}{\partial \theta} \hat{\theta} + \frac{\partial \psi}{\partial z} \hat{z} \\ \nabla \cdot \vec{A} &= \frac{1}{r} \frac{\partial}{\partial r} (r A_r) + \frac{1}{r} \frac{\partial A_\theta}{\partial \theta} + \frac{\partial A_z}{\partial z} \\ \nabla \times \vec{A} &= \hat{r} \left( \frac{1}{r} \frac{\partial A_z}{\partial \theta} - \frac{\partial A_\theta}{\partial z} \right) + \hat{\theta} \left( \frac{\partial A_r}{\partial z} - \frac{\partial A_z}{\partial r} \right) + \frac{1}{r} \hat{z} \left( \frac{\partial}{\partial r} (r A_\theta) - \frac{\partial A_\theta}{\partial r} \right) \\ \nabla^2 \psi &= \frac{1}{r} \frac{\partial}{\partial r} \left( r \frac{\partial \psi}{\partial r} \right) + \frac{1}{r^2} \frac{\partial^2 \psi}{\partial \theta^2} + \frac{\partial^2 \psi}{\partial z^2} \end{aligned}$$

Spherical:

$$\begin{aligned} ds^2 &= dr^2 + r^2 d\theta^2 + r^2 \sin^2 \theta d\phi^2 \\ dV &= r^2 \sin \theta dr d\theta d\phi \\ \nabla \psi &= \frac{\partial \psi}{\partial r} \hat{r} + \frac{1}{r} \frac{\partial \psi}{\partial \theta} \hat{\theta} + \frac{1}{r \sin \theta} \frac{\partial \psi}{\partial \phi} \hat{\phi} \\ \nabla \cdot \vec{A} &= \frac{1}{r^2} \frac{\partial}{\partial r} (r^2 A_r) + \frac{1}{r \sin \theta} \frac{\partial}{\partial \theta} (\sin \theta A_\theta) + \frac{1}{r \sin \theta} \frac{\partial A_\phi}{\partial \phi} \\ \nabla \times \vec{A} &= \frac{1}{r \sin \theta} \hat{r} \left[ \frac{\partial}{\partial \theta} (\sin \theta A_\phi) - \frac{\partial A_\theta}{\partial \phi} \right] + \hat{\theta} \left[ \frac{1}{r \sin \theta} \frac{\partial A_r}{\partial \phi} - \frac{1}{r} \frac{\partial}{\partial r} (r A_\phi) \right] + \frac{1}{r} \hat{\phi} \left[ \frac{\partial}{\partial r} (r A_\theta) - \frac{\partial A_r}{\partial \theta} \right] \\ \nabla^2 \psi &= \frac{1}{r^2} \frac{\partial}{\partial r} \left( r^2 \frac{\partial \psi}{\partial r} \right) + \frac{1}{r^2 \sin \theta} \frac{\partial}{\partial \theta} \left( \sin \theta \frac{\partial \psi}{\partial \theta} \right) + \frac{1}{r^2 \sin^2 \theta} \frac{\partial^2 \psi}{\partial \phi^2} \end{aligned}$$

### References

- [1] J. Murphy, Synchrotron Light Source Data Book, BNL 42333 (version 3.0) (1993)

### 1.5.3 Electromagnetism

$$\begin{aligned} \nabla \cdot \vec{B} &= 0 \\ \nabla \cdot \vec{D} &= \rho \\ \nabla \times \vec{H} - \frac{\partial \vec{D}}{\partial t} &= \vec{J} \\ \nabla \times \vec{E} + \frac{\partial \vec{B}}{\partial t} &= 0 \\ \vec{D} &= \epsilon \vec{E}, \quad \vec{B} = \mu \vec{H} \\ \vec{B} &= \nabla \times \vec{A}, \quad \vec{E} = -\nabla \Phi - \frac{\partial \vec{A}}{\partial t} \\ \text{Continuity} \quad \frac{\partial \rho}{\partial t} + \nabla \cdot \vec{J} &= 0 \\ \text{static Coulomb force} \quad \vec{F} &= \frac{q_1 q_2}{4\pi\epsilon_0 r^2} \hat{r} \\ \text{Lorentz force} \quad \vec{F} &= q(\vec{E} + \vec{v} \times \vec{B}) \\ \text{energy density} \quad u &= \frac{1}{2}(\vec{D} \cdot \vec{E} + \vec{H} \cdot \vec{B}) \\ \text{momentum density} \quad \vec{g} &= \frac{1}{c^2} \vec{E} \times \vec{H} \\ \text{Poynting vector} \quad \vec{S} &= \vec{E} \times \vec{H} \end{aligned}$$

Steady-state boundary conditions between two media:

$$\begin{aligned} (\vec{D}_2 - \vec{D}_1) \cdot \hat{n} &= \rho_{\text{surface}} \\ (\vec{B}_2 - \vec{B}_1) \cdot \hat{n} &= 0 \\ (\vec{E}_2 - \vec{E}_1) \times \hat{n} &= 0 \\ (\vec{H}_2 - \vec{H}_1) \times \hat{n} &= -\vec{J}_{\text{surface}} \end{aligned}$$

where  $\rho_{\text{surface}}$  is the surface charge density;  $\vec{J}_{\text{surface}}$  is the surface current on the boundary;  $\hat{n}$  is the unit vector normal to the boundary and points into medium 2.

### 1.5.4 Kinematical Relations

Relations between  $\beta$ ,  $cp$ , rest mass energy  $E_0$ , kinetic energy  $T$ ,  $E$ , and  $\gamma$  [1]:

|            | $\beta$                                      | $cp$  | $T$   | $E$   | $\gamma$                 |
|------------|--|---|---|---|--------------------------|
| $\beta =$  | $\beta$                                      | $\frac{cp/E_0}{\sqrt{(cp/E_0)^2 + 1}}$  | $\sqrt{1 - (1 + \frac{T}{E_0})^{-2}}$                                       | $\sqrt{1 - (\frac{E_0}{E})^2} = \frac{cp}{E}$ | $\sqrt{1 - \gamma^{-2}}$ |
| $cp =$     | $E_0/\sqrt{\beta^{-2} - 1}$<br>$= E\beta$    | $cp$  | $[T(2E_0 + T)]^{1/2}$<br>$= T \left(\frac{\gamma+1}{\gamma-1}\right)^{1/2}$ | $\sqrt{E^2 - E_0^2}$<br>$= E\beta$            | $E_0\sqrt{\gamma^2 - 1}$ |
| $E_0 =$    | $cp/\beta\gamma$<br>$= E(1 - \beta^2)^{1/2}$ | $cp(\gamma^2 - 1)^{-1/2}$   | $T/(\gamma - 1)$  | $\sqrt{E^2 - c^2 p^2}$                        | $E/\gamma$               |
| $T =$      | $[\frac{1}{\sqrt{1-\beta^2}} - 1]E_0$        | $\sqrt{E_0^2 + c^2 p^2} - E_0$<br>$= cp \left(\frac{\gamma-1}{\gamma+1}\right)^{1/2}$ | $T$   | $E - E_0$                                     | $E_0(\gamma - 1)$        |
| $\gamma =$ | $(1 - \beta^2)^{-1/2}$                       | $\frac{cp}{E_0\beta} = [1 - (\frac{cp}{E_0})^2]^{1/2}$                                | $1 + T/E_0$   | $E/E_0$                                       | $\gamma$                 |

First derivatives:

|                    | $d\beta$   | $d(cp)$  | $d\gamma = dE/E_0 = dT/E_0$  |
|--------------------|--|--|--|
| $d\beta =$         | $d\beta$   | $[1 + (cp/E_0)^2]^{-3/2} d(cp)/E_0$<br>$= \gamma^{-3} d(cp)/E_0$ | $\gamma^{-2}(\gamma^2 - 1)^{-1/2} d\gamma$<br>$= \beta^{-1} \gamma^{-3} d\gamma$ |
| $d(cp) =$          | $E_0(1 - \beta^2)^{-3/2} d\beta = E_0 \gamma^3 d\beta$ | $d(cp)$  | $\frac{E_0 \gamma d\gamma}{\sqrt{\gamma^2 - 1}} = \frac{E_0 d\gamma}{\beta}$     |
| $d\gamma = dE/E_0$ | $\beta(1 - \beta^2)^{-3/2} d\beta$                     | $[1 + (E_0/cp)^2]^{-1/2} d(cp)/E_0$                              | $d\gamma$  |
| $= dT/E_0 =$       | $= \beta \gamma^3 d\beta$                              | $= \beta d(cp)/E_0$  |  |

Logarithmic first derivatives:

|                    | $d\beta/\beta$                    | $dp/p$  | $dT/T$                           | $dE/E = d\gamma/\gamma$   |
|--------------------|-----------------------------------|---|----------------------------------|---|
| $d\beta/\beta =$   | $d\beta/\beta$                    | $\gamma^{-2} dp/p$<br>$= dp/p - d\gamma/\gamma$ | $[\gamma(\gamma + 1)]^{-1} dT/T$ | $(\gamma^2 - 1)^{-1} d\gamma/\gamma$<br>$= (\beta\gamma)^{-2} d\gamma/\gamma$ |
| $dp/p =$           | $\gamma^2 d\beta/\beta$           | $dp/p$  | $[\gamma/(\gamma + 1)] dT/T$     | $\beta^{-2} d\gamma/\gamma$   |
| $dT/T =$           | $\gamma(\gamma + 1) d\beta/\beta$ | $(1 + \gamma^{-1}) dp/p$                        | $dT/T$                           | $\gamma(\gamma - 1)^{-1} d\gamma/\gamma$                                      |
| $dE/E =$           | $(\beta\gamma)^2 d\beta/\beta$    | $\beta^2 dp/p$                                  | $(1 - \gamma^{-1}) dT/T$         | $d\gamma/\gamma$  |
| $d\gamma/\gamma =$ | $= (\gamma^2 - 1) d\beta/\beta$   | $= dp/p - d\beta/\beta$                         |                                  |   |

Relations between bending magnetic field  $B$ , momentum  $p$ , mean orbit radius  $R$ , and revolution frequency  $f$ : Of the four quantities  $(B, p, R, f)$ , only two can be independently chosen, yielding the table on the right ( $\gamma_t$  is the transition gamma).

| Variables   | differential relation  |
|-------------|--|
| $(B, p, R)$ | $\frac{dp}{p} = \gamma_t^2 \frac{dR}{R} + \frac{dB}{B}$  |
| $(f, p, R)$ | $\frac{dp}{p} = \gamma^2 \frac{df}{f} + \gamma^2 \frac{dR}{R}$                                 |
| $(B, f, p)$ | $\frac{dB}{B} = \gamma_t^2 \frac{df}{f} + \frac{\gamma^2 - \gamma_t^2}{\gamma^2} \frac{dp}{p}$ |
| $(B, f, R)$ | $\frac{dB}{B} = \gamma^2 \frac{df}{f} + (\gamma^2 - \gamma_t^2) \frac{dR}{R}$                  |

### References

- [1] C. Bovet et al., CERN/MPS-SI/Int. DL/70/4 (1970)

## Sec.1.6: GLOSSARY OF ACCELERATOR TYPES

### 1.5.5 Vector Analysis

$$\begin{aligned}
 \vec{a} \cdot (\vec{b} \times \vec{c}) &= \vec{b} \cdot (\vec{c} \times \vec{a}) = \vec{c} \cdot (\vec{a} \times \vec{b}) \\
 \vec{a} \times (\vec{b} \times \vec{c}) &= (\vec{a} \cdot \vec{c})\vec{b} - (\vec{a} \cdot \vec{b})\vec{c} \\
 (\vec{a} \times \vec{b}) \cdot (\vec{c} \times \vec{d}) &= (\vec{a} \cdot \vec{c})(\vec{b} \cdot \vec{d}) - (\vec{a} \cdot \vec{d})(\vec{b} \cdot \vec{c}) \\
 \nabla \times \nabla \psi &= 0 \\
 \nabla \cdot (\nabla \times \vec{a}) &= 0 \\
 \nabla(\phi\psi) &= \phi\nabla\psi + \psi\nabla\phi \\
 \nabla \times (\nabla \times \vec{a}) &= \nabla(\nabla \cdot \vec{a}) - \nabla^2 \vec{a} \\
 \nabla \cdot (\psi \vec{a}) &= \vec{a} \cdot \nabla \psi + \psi \nabla \cdot \vec{a} \\
 \nabla \times (\psi \vec{a}) &= \nabla \psi \times \vec{a} + \psi \nabla \times \vec{a} \\
 \nabla(\vec{a} \cdot \vec{b}) &= (\vec{a} \cdot \nabla)\vec{b} + (\vec{b} \cdot \nabla)\vec{a} \\
 &\quad + \vec{a} \times (\nabla \times \vec{b}) + \vec{b} \times (\nabla \times \vec{a}) \\
 \nabla \cdot (\vec{a} \times \vec{b}) &= \vec{b} \cdot (\nabla \times \vec{a}) - \vec{a} \cdot (\nabla \times \vec{b}) \\
 \nabla \times (\vec{a} \times \vec{b}) &= \vec{a}(\nabla \cdot \vec{b}) - \vec{b}(\nabla \cdot \vec{a}) \\
 &\quad + (\vec{b} \cdot \nabla)\vec{a} - (\vec{a} \cdot \nabla)\vec{b} \\
 \vec{a} \times (\vec{b} \times \vec{c}) + \vec{b} \times (\vec{c} \times \vec{a}) + \vec{c} \times (\vec{a} \times \vec{b}) &= 0
 \end{aligned}$$

### 1.5.6 Relativity

Let  $F$  be the stationary laboratory frame with space time coordinates  $(\vec{x}, t)$ . Let  $F'$  with  $(\vec{x}', t')$  be a frame moving with velocity  $\vec{V} = c\vec{\beta}$  with respect to  $F$ . Lorentz transformations:

Coordinates :

$$\vec{x}' = \vec{x} + \gamma \vec{\beta} \left( \frac{\gamma}{\gamma+1} \vec{\beta} \cdot \vec{x} - ct \right)$$

$$t' = \gamma \left( t - \frac{1}{c} \vec{\beta} \cdot \vec{x} \right)$$

Velocity :  $\vec{v}' = \frac{\vec{v} + \gamma \vec{V} \left( \frac{\gamma}{\gamma+1} \frac{\vec{v} \cdot \vec{V}}{c^2} - 1 \right)}{\gamma \left( 1 - \frac{\vec{v} \cdot \vec{V}}{c^2} \right)}$

Energy-momentum :

$$\vec{P}' = \vec{P} + \gamma \vec{\beta} \left( \frac{\gamma}{\gamma+1} \vec{\beta} \cdot \vec{P} - \frac{1}{c} E \right)$$

$$E' = \gamma(E - c\vec{\beta} \cdot \vec{P})$$

EM fields :

$$\vec{E}' = \gamma(\vec{E} + c\vec{\beta} \times \vec{B}) - \frac{\gamma^2}{\gamma+1} (\vec{\beta} \cdot \vec{E}) \vec{\beta}$$

$$\vec{B}' = \gamma(\vec{B} - \frac{1}{c} \vec{\beta} \times \vec{E}) - \frac{\gamma^2}{\gamma+1} (\vec{\beta} \cdot \vec{B}) \vec{\beta}$$

When  $\vec{V} = V\hat{x}$ , the above becomes

Coordinates :

$$x' = \gamma(x - Vt), \quad t' = \gamma \left( t - \frac{Vx}{c^2} \right)$$

$$y' = y, \quad z' = z$$

Velocity :  $v'_x = \frac{v_x - V}{1 - \frac{\beta v_x}{c}}$

$$v'_y = \frac{v_y / \gamma}{1 - \frac{\beta v_x}{c}}, \quad v'_z = \frac{v_z / \gamma}{1 - \frac{\beta v_x}{c}}$$

Energy-momentum :

$$P'_x = \gamma(P_x - \frac{\beta}{c} E), \quad E' = \gamma(E - c\beta P_x)$$

$$P'_y = P_y, \quad P'_z = P_z$$

EM fields :

$$E'_x = E_x, \quad B'_x = B_x$$

$$E'_y = \gamma(E_y - c\beta B_z), \quad B'_y = \gamma(B_y + \frac{\beta}{c} E_z)$$

$$E'_z = \gamma(E_z + c\beta B_y), \quad B'_z = \gamma(B_z - \frac{\beta}{c} E_y)$$

## 1.6 GLOSSARY OF ACCELERATOR TYPES

### 1.6.1 Antiproton Sources

*K. Gollwitzer, J. Marriner, FNAL*

Antiproton ( $\bar{p}$ ) sources are complete accelerator complexes utilizing many accelerator technologies [1]–[4]. A primary proton beam is used to produce  $\bar{p}$ 's on a target. The production process is inefficient, and the secondary  $\bar{p}$  beam is several orders of magnitude less dense than the primary proton beam. The design strategy consists of maximizing the  $\bar{p}$  phase space density by (a) appropriate preparation of the proton beam and (b) beam cooling using some combination of stochastic cooling (Sec.2.7.1) and electron cooling (Sec.2.7.2).

It is advantageous for the proton beam to have a short time spread and a small spot size at the  $\bar{p}$  target since these properties are transferred to the secondary  $\bar{p}$  beam. The proton beam can obtain a short bunch length by a bunch rotation process (Sec.4.10). The rf voltage is lowered so that the beam fills the bucket and then raised to its maximum value. The bunch rotates  $90^\circ$  in phase space and achieves a momentary short length. The proton beam is strongly focused transversely to a minimum spot size at the target; see Fig.1. The resulting beam focal length is comparable to typical nuclear interaction lengths, favoring the use of high- $Z$  targets, but excessive energy deposition from pair production limits their usefulness. Tab.1 shows typical proton beam parameters.

The production cross-section has been parameterized as [5]

$$\begin{aligned}
 \frac{E}{\sigma_{\text{abs}}} \frac{d^3\sigma}{dp_t^3} &= [0.065(1 - x_r)^8 \exp(-3p_t^2)] \\
 &\quad \times [1 + 24s^{-2} \exp(8x_r)] \\
 &\quad \times [a \exp(bp_t^2) \exp(-cx_r)]
 \end{aligned}$$

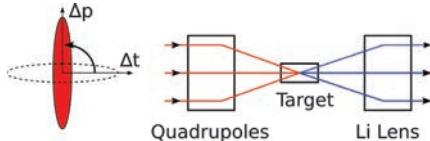


Figure 1: Antiproton Targeting. The proton beam is given a short time spread via bunch rotation and focused to a small spot size on the production target. The secondary beam is focused by a lithium lens or horn.

Table 1: Proton beam parameters for the FNAL Main Injector and CERN Proton Synchrotron.

| Parameter                              | FNAL<br>(MI) | CERN<br>(PS) |
|--|--------------|--------------|
| Momentum (GeV/c)                       | 120          | 26           |
| Protons/pulse ( $10^{12}$ )            | 8            | 15           |
| Cycle time (s)                         | 2.2          | 100          |
| # Bunches                              | 81           | 5            |
| RF bucket length (ns)                  | 19           | 105          |
| RF bunch length (ns)                   | 1.5          | 5            |
| Beam radius $\sigma$<br>at target (mm) | 0.16         | 1.5          |

with  $\sigma_{\text{abs}}$  the absorption cross-section,  $p_t$  the transverse momentum [GeV/c],  $x_r$  the radial scaling variable ( $= E/E_{\max}$ ,  $\bar{p}$  energy in CM divided by its kinematically allowed value),  $s$  the square of the CM energy [GeV $^2$ ]. The parameterization consists of three multiplicative factors: (i) a cross section for a hydrogen target at infinite energy, (ii) deviations from scaling (including  $s$ -dependence), and (iii) the nuclear dependence. Values for  $a, b, c$  for various nuclei are given in [5]. For copper,  $a = 1.50$ ,  $b = 1.43$ ,  $c = 1.56$ .

The initial large angular divergence of the  $\bar{p}$  beam is largely eliminated by a collection element: a high-gradient lithium lens (Sec.7.2.12) or a pulsed horn. The large momentum spread is reduced as the beam is debunched through a second bunch rotation. The beam is then pre-cooled with stochastic cooling. The parameters of the collected beam are given in Tab.2.

A difficult step is to accumulate  $\bar{p}$ 's from  $10^4$  or more pulses. The stacking process is accomplished with stochastic cooling in an accumulator ring; see Fig.2. The flux of  $\bar{p}$ 's that can be accumulated is [6]

$$\Phi_0 = \frac{E_d T_0 W^2 |\eta|}{\beta^2 E \Lambda} \quad (1)$$

Table 2: Beam parameters of the collected  $\bar{p}$  beam.  $6\pi\epsilon$  is for 95% beam, un-normalized.  $\Delta p/p$  is the full momentum spread of the beam.

| Parameter                             | FNAL<br>(Debunch.) | CERN<br>(AD) |
|---------------------------------------|--------------------|--------------|
| Momentum (GeV/c)                      | 8.9                | 3.57         |
| $\Delta p/p$ (%)                      |                    |              |
| Before rotation                       | 4.5                | 6            |
| After rotation                        | 0.2                | 1.5          |
| After cooling                         | 0.1                | 0.18         |
| $6\pi\epsilon$ (mm-mrad)              |                    |              |
| Before cooling                        | $35\pi$            | $200\pi$     |
| After cooling                         | $3\pi$             | $5\pi$       |
| Stochastic cooling<br>bandwidth (GHz) | 4-8                | 0.9-1.6      |

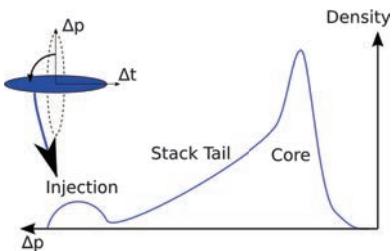


Figure 2: Antiproton Stacking. A small momentum spread is recovered from the secondary beam via a bunch rotation and the beam is injected into an accumulation ring where the longitudinal density is increased by stochastic stacking. Individual particles are stochastically decelerated towards the stack core, where high density beams can be formed and extracted.

where  $W = f_{\max} - f_{\min}$  is the system bandwidth,  $\beta$  is the relativistic velocity factor,  $E$  is the beam energy,  $\eta$  = slip factor,  $\Lambda = \ln(f_{\max}/f_{\min})$ , and  $\frac{1}{E_d} = \frac{1}{V} \frac{dV}{dE}$  with  $V(E)$  the voltage gain per turn. In [6],  $V$  is exponentially decreasing with  $E$  while the particle density increases as a function of energy  $\Psi(E) \propto e^{E/E_d}$ . For the FNAL Accumulator,  $\eta = 0.02$ ,  $E_d \sim 10$  MeV,  $E = 8.9$  GeV. Some parameters of the stacking process are given in Tab.3.

At FNAL, the  $\bar{p}$  beam is transferred from the Accumulator to the Recycler, a large storage ring that stacks the beam azimuthally using barrier bucket rf manipulations (Sec.4.10). Both stochastic and electron cooling are used to increase the  $\bar{p}$  beam density. The electron cooling system is notable for its high energy electron beam (4.9 MeV). Final  $\bar{p}$  beam parameters are in Tab.4.

## Sec.1.6: GLOSSARY OF ACCELERATOR TYPES

Table 3: Beam properties in the accumulation ring.

| Parameter                                | FNAL<br>(Accum.) |
|--|------------------|
| Stack Rate ( $10^{10}$ hr $^{-1}$ )      | 28               |
| $\bar{p}$ /pulse ( $10^7$ )              | 20               |
| Yield ( $\bar{p}$ /proton) ( $10^{-6}$ ) | 24               |
| Final $\Delta p/p$ (%)                   | 0.2              |
| Final emit. $6\pi\epsilon$<br>(mm-mrad)  | $1.5\pi$         |
| Bandwidths (GHz)                         |                  |
| Stacktail                                | 2-4              |
| Core cooling                             | 2-4,4-8          |
| Total particles ( $10^{12}$ )            |                  |
| typical, max                             | 0.25, 3          |

Table 4: Final antiproton beam properties.

| Parameter                               | FNAL<br>(Recyc.)     | CERN<br>(AD)    |
|---|----------------------|-----------------|
| Cycle time                              | 20 hr                | 100 s           |
| Total particles                         | $5.5 \times 10^{12}$ | $1 \times 10^7$ |
| Final $\Delta p/p$ (%)                  | 0.04                 | 0.01            |
| Final emit. $6\pi\epsilon$<br>(mm-mrad) | $0.2\pi$             | $1\pi$          |

At CERN the  $\bar{p}$  source consisted originally of an antiproton accumulator (AA) to which was added the antiproton collector (AC) to increase production by increasing the  $\bar{p}$  beam acceptance and by using the bunch rotation technique pioneered at FNAL. The low energy antiproton ring (LEAR) was the first accelerator dedicated to particle experiments with  $\bar{p}$ 's. LEAR has been decommissioned, but low energy  $\bar{p}$ 's continue to be produced by the CERN Antiproton Decelerator (AD), a reconfiguration of the original CERN  $\bar{p}$  source that emphasizes economical production of low energy  $\bar{p}$ 's [4]. Antiprotons are not accumulated but each pulse of  $\bar{p}$ 's is decelerated and cooled several times before extracting the entire beam for low energy/stopping beam experiments. Stochastic cooling is used initially while electron cooling is used at the lower energies. Tab.4 gives typical beam parameters for the current CERN and FNAL sources, which have now significantly diverged in their primary uses.

The FAIR project at GSI [7] will include a  $\bar{p}$  source and storage rings to support experiments with  $\bar{p}$  beams. FAIR will provide high energy  $\bar{p}$  beams (up to 14 GeV) interacting with an internal gas jet target at high luminosity ( $2 \times 10^{32} \text{ cm}^{-2}\text{s}^{-1}$ ), extending the technique that was first used briefly in the CERN ISR and then later

with the FNAL Accumulator. FAIR will also provide low energy  $\bar{p}$  beams (down to 30 MeV) for experiments with low energy or stopped  $\bar{p}$ 's.

Antiproton beams have been used to produce relativistic antihydrogen atoms at CERN [8] and Fermilab [9]. The relativistic antiatoms lasted a few tens of ns. Experiments at the AD have used trapped  $\bar{p}$ 's to produce antihydrogen [10, 11]; with the antihydrogen existing long enough for physics measurements [12].  $\bar{p}$ 's will be used at the AD and FAIR in the future to study antihydrogen.

## References

- [1] M.D. Church, J. Marriner, Ann. Rev. Nucl. Part. Sci. 43 (1993) 253
- [2] B. Autin et al, CERN/PS/AA 78-3 (1978); E.J. Wilson et al, CERN 83-10 (1983)
- [3] Fermilab Design Report, Tevatron I Project (1984)
- [4] S. Maury, ed., CERN/PS 96-43 (1996)
- [5] C. Hojvat, A. van Ginneken, NIM 206 (1983) 67
- [6] S. van der Meer, CERN/PS/AA 78-22 (1978)
- [7] H.H. Gutbrod, ed., FAIR Baseline Technical Report (2006), <http://www.gsi.de/fair/reports/btr.html>
- [8] G. Bauer et al, Phys. Lett. B 368 (1996) 251
- [9] G. Blanford et al, PRL 80 (1998) 3037
- [10] M. Moretti et al, Nature 419 (2002) 456
- [11] G. Gabrielese et al, PRL 89 (2002) 233401
- [12] G.B. Andresen et al, Nature Physics 7 (2011)

### 1.6.2 Betatron

*M. Tigner, Cornell U.*

The Betatron [1] is a cyclic electron accelerator with a circular orbit of approximately constant radius which provides acceleration through Magnetic Induction. Iron core machines of energies up to 300 MeV have been constructed. The guide field in the classical Betatron is weak focusing. A section perpendicular to the equilibrium orbit plane is shown in Fig.1 [2].

The beam travels in a doughnut shaped, evacuated, dielectric (e.g. glass) chamber with a thin conducting film on the inside to prevent charging. The accelerating electric field is produced by the changing magnetic flux within the equilibrium orbit. By combining Newton's law and Faraday's law of induction together with the Lorentz force law we can write

$$\frac{dp}{dt} = \frac{e}{2\pi\rho} \frac{\partial\Phi}{\partial t}, \quad p = Be\rho \quad (1)$$

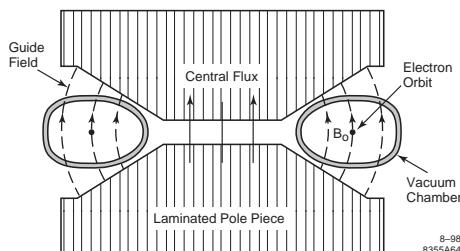


Figure 1: Betatron schematic.

and find that

$$\Delta\Phi = 2\pi\rho^2 B \quad (2)$$

the famous 2 to 1 condition that the flux change within the orbit must be twice that which one would obtain if the field were uniform throughout the region inside the orbit and equal to the field at the orbit.

Since the field in an iron core magnet is limited by saturation, increase in maximum energy means increase in radius of the core and thus the volume of the iron. The iron of the 300 MeV Illinois Betatron weighed > 300 tons.

There is another constraint on the maximum energy which can be obtained using the betatron principle. As the beam energy rises synchrotron radiation loss rises and competes with the energy gain due to magnetic induction. This effect spoils the linearity of Eq.(1) and requires special means for adding the extra energy needed. In practice, the synchrotron radiation begins to become important at  $\sim 100$  MeV and limits beam energies that can be obtained with iron and copper magnet technology to  $\sim 300$  MeV.

While betatrons using the combined function, one magnetic circuit, design of Fig.1 have been successfully operated at low energies, considerable efficiency in size and operation can be achieved by separating the functions as shown in Fig.2 [2]. The scale refers to an 80 MeV machine. The magnetic elements are punched 0.355 mm laminations of silicon transformer steel, wound with litz wire to minimize eddy current. Even so the highest energy betatrons achieve only about 5% duty factor due to the large hysteresis and eddy current losses coupled with the difficulty of cooling the massive iron cores. Injection is accomplished by insertion of an electron gun into the sealed, doughnut vacuum chamber, just outside or just inside the equilibrium orbit, and pulsing it negatively at tens to more than 100 kV for

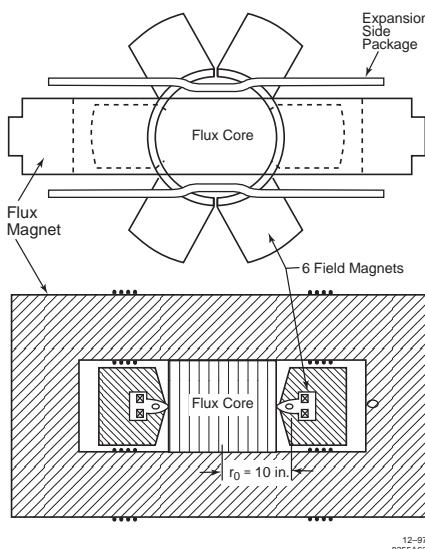


Figure 2: Separated function betatron.

a few microseconds at injection time. Fig.3 [2] shows some typical power cycles of various betatrons. Injection is complicated and involves space charge effects (Sec.2.4.5.1) in a central way [3]. Injection efficiency can be greatly enhanced by adding pulsed coils to draw the equilibrium orbit temporarily away from the gun just after injection and slowly restoring it as the initial betatron oscillations damp. Ejection or targeting is done in the same way by pulsewise distortion of the orbit enough to drive the beam into a weak field region and out of the doughnut through a thin window or into a tungsten bremsstrahlung target within the doughnut with subsequent extraction of the x-rays through a thin window.

While some betatrons, particularly the 300 MeV machine were used for nuclear physics research, the majority were used for medical therapy or diagnostic x-raying of industrial equipment. The 300 MeV machine was capable of producing 14,600 R/min. at 1 m in Pb while the machines used for therapy produce typically about 100 R/min. at 1 m. These machines have been largely supplanted by linacs which are more powerful, more flexible, lighter and more easily controlled. In an attempt to overcome the space charge engendered limitation to beam current in the classical betatron, the concept of the "Modified Betatron" was developed. In the Modified Betatron, a toroidal field and strong

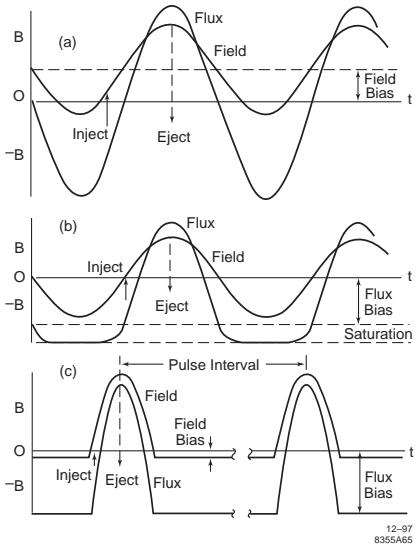


Figure 3: Typical power cycles.

focusing both are added to the normal, weak focusing, vertical (mirror) field of the classical betatron. A beam of  $\sim 1$  kA has been accelerated this way to 20 MeV [4]. Other new ideas have come forward recently, inspired by new materials and new needs. One approach proposes very high frequency operation to achieve high beam power [5]. Another approach uses the FFAG principle (Sec.1.6.6) in a betatron configuration [6].

## References

- [1] M.S. Livingston, J.P. Blewett, Particle Accelerators, McGraw-Hill (1962)
- [2] By permission, McGraw-Hill Book Co.
- [3] L. Gonella, Supplement to Nuovo Cimento 3 (1966) 303
- [4] C.A. Kapetanakos et al, Phys. Fluids B5 (1993) 2295
- [5] G.V. Dolbilov et al, APAC 2007 (2007) 628
- [6] S. Boucher et al, EPAC08 (2008) 1860

### 1.6.3 Colliders

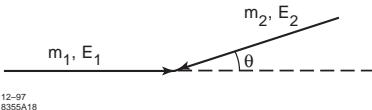
*D. Hartill, Cornell U.*

Particle beam colliders were developed to increase the center of mass energy available for new particle production and interaction. For a stationary target this energy scales only as the square root of the beam energy while with a collider the collision energy scales linearly allowing for much

higher available energies. The center of mass energy  $E_{cm}$  for the collision of two particles of mass  $m_1$  and  $m_2$  with energies  $E_1$  and  $E_2$  with a crossing angle  $\theta$  is

$$E_{cm} = \left[ 2E_1E_2 + (m_1^2 + m_2^2)c^4 \right]^{1/2} \quad (1)$$

$$+ 2(E_1^2 - m_1^2c^4)^{1/2}(E_2^2 - m_2^2c^4)^{1/2} \cos \theta \right]^{\frac{1}{2}}$$



Typically  $m_1 = m_2$  and  $E_1 = E_2 (= E)$  and  $\theta = 0$  with  $E_{cm} = 2E$  and the CM is stationary in the laboratory. With the advent of B factories, beams of  $e^-$ 's and  $e^+$ 's are collided with different energies so that the CM is moving with high velocity. Short lived particles produced in the collisions then move a measurable distance before decay enabling important experiments in CP violation. In this case,  $E_{cm} \approx 2(E_1E_2)^{1/2}$  for  $\theta = 0$ .

Colliding beam storage rings were first described in 1956 [1]. Princeton-Stanford 500 MeV two ring  $e^-e^-$ , the two ring 140 MeV  $e^-e^-$  ring VEP-1 at Novosibirsk, and the 250 MeV  $e^+e^-$  AdA at Frascati were the first operating  $e^\pm$  storage rings. The ISR at CERN was the first hadron collider that collided protons using two intersecting storage rings [2]. The first linear collider was the SLAC Linear Collider (SLC) and was based on an idea that was first proposed in 1965 [3]. The SLC used a single linac with two beamline arcs to collide the particles.

Most colliders are based on storage rings where the beams are allowed to interact at one or more collision points and can be either double rings or a single ring employing electrostatic separation to keep the oppositely charged beams apart except at the collision points. The beam particles can interact many times in this configuration. Another configuration is the use of one or two linacs with the beam particles colliding only once. The proposed ILC [4] and CLIC [5] colliders use two linacs each aimed at the collision point.

The interaction rate of a collider is measured by its luminosity (Sec.4.1) with typical units of  $10^{32} \text{ cm}^{-2}\text{s}^{-1}$ . Various instabilities can limit the performance of a collider but a variety of feedback systems and careful selection of operating

Table 1: Colliding Beam Machines. [a] DR: Double storage ring. SR: Single storage ring. LC: Linear collider.  
 [b] Princeton-Stanford Colliding Beam Experiment. [c] 200 GeV per charge unit. [d] Planned.

| Location             | Name (type <sup>[a]</sup> )       | Beam Energies $E$ (GeV)                     | Start               |
|----------------------|-----------------------------------|---|---------------------|
| Stanford/SLAC, USA   | CBX <sup>[b]</sup> ( $e^-e^-$ DR) | .5 + .5                                     | 1963                |
|                      | Spear ( $e^+e^-$ SR)              | 2.5 + 2.5                                   | 1972                |
|                      | PEP ( $e^+e^-$ SR)                | 15 + 15                                     | 1980                |
|                      | SLC ( $e^+e^-$ LC)                | 50 + 50                                     | 1989                |
|                      | PEP-II ( $e^+e^-$ DR)             | 3.1 + 9.0                                   | 1999                |
| Frascati, Italy      | AdA ( $e^+e^-$ SR)                | 0.25 + 0.25                                 | 1962                |
|                      | Adone ( $e^+e^-$ SR)              | 1.5 + 1.5                                   | 1969                |
|                      | DAΦNE ( $e^+e^-$ SR)              | 0.5 + 0.5                                   | 1997                |
|                      | Super B ( $e^+e^-$ DR)            | 3.1 + 9.0                                   | 2015 <sup>[d]</sup> |
| Novosibirsk, Siberia | VEP-1 ( $e^-e^-$ DR)              | 0.13 + 0.13                                 | 1963                |
|                      | VEPP-2/2M ( $e^+e^-$ SR)          | 0.7 + 0.7                                   | 1974                |
|                      | VEPP-4 ( $e^+e^-$ SR)             | 7 + 7                                       | 1979                |
| Cambridge, USA       | CEA Bypass ( $e^+e^-$ SR)         | 3.0 + 3.0                                   | 1971                |
| Orsay, France        | ACO ( $e^+e^-$ SR)                | 0.5 + 0.5                                   | 1966                |
| DESY, Germany        | DCI ( $e^+e^-$ DR)                | 1.8 + 1.8                                   | 1976                |
|                      | Doris ( $e^+e^-$ SR)              | 3 + 3                                       | 1974                |
|                      | Petra ( $e^+e^-$ SR)              | 19 + 19                                     | 1978                |
|                      | HERA ( $e^\pm p$ DR)              | 30 (e) + 820 (p)                            | 1992                |
| CERN, Europe         | ISR (pp DR)                       | 31.5 + 31.5                                 | 1971                |
|                      | S $p\bar{p}$ S (p $\bar{p}$ SR)   | 315 + 315                                   | 1981                |
|                      | LEP ( $e^+e^-$ SR)                | 104.5 + 104.5                               | 1989                |
|                      | LHC (pp, PbPb DR)                 | 7,000 + 7,000                               | 2009                |
| Brookhaven, USA      | RHIC (heavy ion, pp DR)           | 100/u <sup>[c]</sup> + 100/u <sup>[c]</sup> | 1999                |
| Cornell, USA         | CESR ( $e^+e^-$ SR)               | 6 + 6                                       | 1979                |
| KEK, Japan           | Tristan ( $e^+e^-$ SR)            | 30 + 30                                     | 1986                |
| Beijing, China       | KEK B ( $e^+e^-$ DR)              | 3.5 + 8                                     | 1999                |
|                      | Super KEK B ( $e^+e^-$ DR)        | 3.5 + 8                                     | 2015 <sup>[d]</sup> |
|                      | BEPC ( $e^+e^-$ SR)               | 1.55 + 1.55                                 | 1989                |
|                      | BEPC II ( $e^+e^-$ DR)            | 2.1 + 2.1                                   | 2008                |
| Fermilab, USA        | Tevatron (pp-bar SR)              | 980 + 980                                   | 1987                |

points have enabled performance levels in excess of 200 times this unit at the asymmetric B factory at KEK with similar performance at the asymmetric B factory at SLAC. An upgraded Super KEK B [6] at KEK and a new Super B collider [7] to be located near Frascati using parts from the decommissioned SLAC B-factory both project performance levels at 8,000 to 10,000 times this unit. Both projects have been approved by their funding agencies and will take advantage of very low emittance beams made possible by the damping ring technology developed for the SLC. The expected start of commissioning is in the 2014–2016 time frame.

The LHC has begun operation with an initial beam energy of 3.5 TeV and has already achieved a peak luminosity in excess of  $2 \times 10^{32} \text{ cm}^{-2}\text{s}^{-1}$

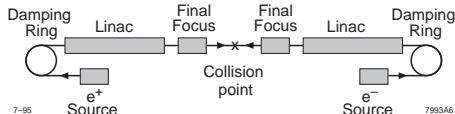


Figure 1: Components of a Linear Collider.

and projects an integrated luminosity of more than  $6 \text{ fb}^{-1}$  over the next two years. The Tevatron collider with a beam energy of 1 TeV has accumulated over  $8 \text{ fb}^{-1}$  per interaction region by the end of its operation in FY2011. The Tevatron collider has been in operation over 23 years and its current peak luminosity of  $4 \times 10^{32} \text{ cm}^{-2}\text{s}^{-1}$  is 400 times its design. The LHC collides protons on protons using two separate rings and the Tevatron collides antiprotons on protons in a single ring

## Sec.1.6: GLOSSARY OF ACCELERATOR TYPES

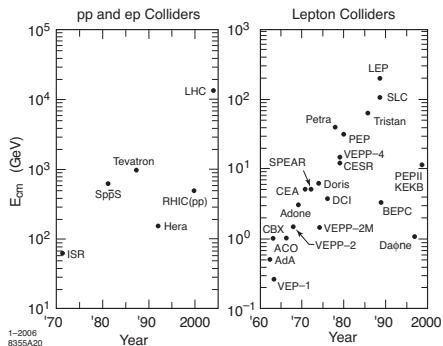


Figure 2: Colliders over the years.

using electrostatic separators to provide separation of the beams except at the two IPs.

Future colliders are likely to be either linear colliders using  $e^-$  and  $e^+$  to avoid high synchrotron radiation losses or storage rings using muons to get to the highest collision energies. Electrons and muons are constituent particles while only the quarks in hadrons are the constituent particles with only a fraction of the total hadron energy. It is the collision energy of the constituent particles that can access new particles.

## References

- [1] D. Kerst et al, PR 102 (1956) 590; G.K. O'Neill, PR 102 (1956) 1418
- [2] K. Johnsen, Proc. 8th Int. Conf. on High Energy Acc. (1971) p.79; CERN AR/Int. SG/64-9 Design Study of ISR (1964)
- [3] M. Tigner, Nuovo Cimento 37 (1965) 1228
- [4] ILC Reference Design Report: [http://media.linearcollider.org/rdr/draft\\_v1.pdf](http://media.linearcollider.org/rdr/draft_v1.pdf) (2007)
- [5] CLIC Design Report: CERN-OPEN-2008-021 (CLIC-Note-764) (2008)
- [6] T. Kageyama, Super KEKB, PANIC 05 Proceedings, p.589 (2005)
- [7] Super B Conceptual Design Report (2007): <http://www.pi.infn.it/SuperB/CDR>

### 1.6.4 Cyclotron

*H. Blosser, Michigan St. U.  
M. Craddock, U. British Columbia &  
TRIUMF*

Earliest [1] and most numerous of circular accelerators, cyclotrons are characterized by magnetic field and accelerating rf frequency which are constant in time (c.w.). (Synrocyclotrons are exceptions. See later.)

Cyclotrons are often referred to by the diameter of the magnet pole (e.g. 27 inch, 184", 2 meter, U-400). More recently "K", ( $\approx$  proton kinetic energy in MeV) has become a designation (e.g. K500, K1200) particularly for multi-particle cyclotrons where the energy for an ion of charge  $Qe$  and mass  $Am_0$  (where  $m_0$  is 1/12 of mass of  $^{12}\text{C}$ ,  $m_0c^2 = 931.48$  MeV) is given nonrelativistically by  $E = KQ^2/A$ . The maximum bending power ( $B\rho$ ) is related to  $K$  by  $K = (eB\rho)^2/(2m_0)$ . Many cyclotrons are referred to by a local name or acronym (e.g. ORIC, AGOR). Typical beam parameters achieved by cyclotrons: normalized emittances  $\epsilon_{Nr} \approx 2$  to 0.2 mm-mrad,  $\epsilon_{Nz} \approx 5$  to 1 mm-mrad, and energy spread  $\Delta E/E$  typically  $10^{-3}$  with best value  $2 \times 10^{-4}$ .

Cyclotrons have evolved in many, sometimes overlapping, subclassifications. Their characteristics are best documented in the proceedings of a series of triennial conferences [2], and are summarized in [3].

**Classical cyclotron (now rare)** Fig.1 shows the original cyclotron concept [1]. A static and uniform magnetic field  $B$  is applied perpendicular to D-shaped hollow electrodes ("dees"). The dees are driven by an rf voltage whose frequency matches the constant *cyclotron frequency*

$$f = \frac{qB}{2\pi m} \quad (1)$$

of nonrelativistic ions. Ions from a central ion source are repetitively accelerated in and out of the dees on a spiral path to maximum energy. In practice, the magnetic field must decrease with  $r$  to assure stability in the axial direction, and the accelerating voltage must be high for ions to reach the design energy before they get out of the accelerating phase of the rf cycle (due to relativistic mass increase and to magnetic field decreasing with  $r$ ). These effects limit highest ion velocity. The record for classical cyclotrons is  $\beta = 0.22$  (Oak Ridge 86", Stockholm 225-cm).

**Isochronous cyclotron** Thomas [4] pointed out that magnets with alternate strong and weak azimuthal regions ("sectors" or "hills and valleys") provided an additional axial focusing which could offset the defocusing from a radially increasing magnetic field. The average magnetic field can therefore match the mass increase of the accelerated particle with positive axial focusing provided by the azimuthal variation. Two electron cyclotrons with sinusoidal azimuthal variation of the magnetic field were built (Berkeley, 1950,

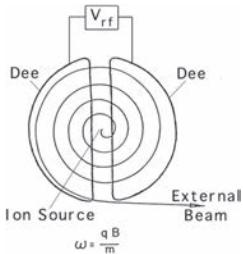


Figure 1: A classical cyclotron. If magnetic field points into page, spiral line is trajectory of positive ion.

$\beta = 0.5$ ). In the 1950s, pole tips with a constant gap in the hill region and a larger constant gap in the valleys came into use, providing a more trapezoidal azimuthal field variation. Also, at that time, “strong-focusing” by spiralling the hills was introduced and designations “sector-focused cyclotron”, “spiral-ridged cyclotron”, “azimuthally-varying-field cyclotron” are now used largely interchangeably with isochronous cyclotron. (Thomas cyclotron is normally reserved for azimuthally sinusoidal fields.) More than 800 isochronous cyclotrons have been built (50-590 MeV for nuclear and particle physics, ~230 MeV for proton therapy, and 10-100 MeV for commercial production of radionuclides). Over a dozen 210-ton IBA 230-MeV cyclotrons have been installed in hospitals worldwide in recent years (Sec.1.6.14).

**Separated sector cyclotron** This is a subclass of isochronous cyclotrons in which the valley regions are iron free. The concept was proposed in the late 1950s by Oak Ridge for a 900 MeV isochronous cyclotron, and first used at PSI in the 590-MeV meson factory (operation 1974). A simpler, “radial-sector” (i.e. non-spiral) formulation of this concept was adopted for the Indiana U. 200 MeV proton cyclotron (operation 1975). The radial, separated sector design is also used in large heavy-ion cyclotrons (GANIL, RIKEN, Lanzhou). The RIKEN K2600 SRC, using superconducting magnet coils, delivers 80-GeV heavy ions, making it the highest energy c.w. accelerator of any type.

**H<sup>-</sup> cyclotron** Cyclotrons to accelerate H<sup>-</sup> ions were introduced in 1962 at U. Colorado [5]. At maximum energy a thin foil strips the electrons from the H<sup>-</sup> ion, reversing their bending radius and thus sending the beam quickly out of the cyclotron. Moving the foil to a different radius easily changes the output energy of such a cyclotron

over a broad range, and difficult extraction components (electrostatic deflectors, magnetic channels) are avoided. Multiple beams can be extracted simultaneously at different energies using suitably shaped foils. The H<sup>-</sup> technique is used in most radionuclide-production cyclotrons, where high beam currents make extraction especially difficult. To avoid electric dissociation (Lorentz stripping, Sec.7.1.8) of H<sup>-</sup> at high velocities, the 520-MeV meson-factory cyclotron at TRIUMF requires weak magnetic fields so that this cyclotron is the world’s largest (17 m dia. poles).

**Superconducting compact cyclotron** This class includes cyclotrons with superconducting main magnet coils or rf systems (e.g. TRITRON at Munich). These cyclotrons can be up to  $\times 10$  lighter than room-temperature cyclotrons due to the unusual cyclotron scaling law that total flux  $\propto 1/B$ . Weight and size reduction leads to  $\sim \times 0.5$  construction cost and  $\sim \times 0.9$  operating costs. Examples: (1) 22-ton cancer therapy cyclotron in Detroit produces 50 MeV deuterons, (2) 240-ton K1200 nuclear physics cyclotron at East Lansing produces beams  $> 10$  GeV for heavier ions, (3) 90-ton Varian/Accel cancer therapy cyclotrons at PSI and Munich produce 250-MeV protons.

**Synchrocyclotron** This largely outmoded form of the classical cyclotron uses an rf frequency which varies with time to track the orbital frequency. The energy limit of the classical cyclotron is relieved, but the c.w. characteristic is lost and beam current typically decreases by  $\times 1000$ . Auxiliary slow extraction systems were often used to stretch the duty cycle. Synchrocyclotrons provided the energy-frontier beams of the 1950s but have now been superseded by synchrotrons for high-energy physics and by isochronous cyclotrons for nuclear physics. The highest beam velocity achieved by a cyclotron is in the 1 GeV synchrocyclotron at Gatchina. A very compact synchrocyclotron for cancer therapy with 250-MeV protons has been built by Still River Systems. The 9-T superconducting magnet is light enough to be mounted on a rotating gantry.

#### Orbit characteristics Magnetic Field

$$B(r, \theta) = B_0(r) + \sum_i B_i(r) \cos[i\theta + \zeta_i(r)] \quad (2)$$

Near  $r = 0$ ,  $B_i(r) \approx (r/G)^i$  where  $G$  is magnet gap.  $B_0 = B_0(r=0) = 2\pi f_0 m_0/q$  [Eq.(1)].

Choice of sector number  $N = 3, 4, 6, 8, \dots$  depends on maximum  $E/A$  (energy/nucleon).  $N = 2$  is radially unstable. Odd numbers other than 3 are too complicated.  $N = 3$  is the most frequent choice below 200 MeV/A because of the fast rise of  $B_3$  near  $r = 0$ .

Flutter is defined as

$$F(r) = \frac{\langle B^2 \rangle - \langle B \rangle^2}{\langle B \rangle^2} = \frac{1}{2} \sum_i f_i(r)^2 \quad (3)$$

where  $f_i(r) = B_i(r)/B_0(r)$ . Sometimes  $f$  is called flutter,  $F$  (sometimes written  $F^2$ ) called mean square flutter. With flutter, orbit scallops,

$$r(\theta) = \langle r \rangle \left[ 1 + \sum_{i \neq 1} g_i \cos(i\theta + \zeta_i) \right] \quad (4)$$

where  $g_i \approx f_i/(i^2 - 1)$ .

Field spiral  $\zeta_N(r)$  is the angular location of maximum value of the main flutter component. Spiral angle  $\alpha$  is the angle between radius vector and tangent line to spiral curve,  $\tan \alpha = rd\zeta/dr$ .

The acceleration system consists of dees (2 gap drift tube) or cavities (1 gap drift tube), now usually located in valleys of magnet (in all valleys or, if  $N$  is even, in alternate valleys). Dees usually operate at 50-200 kV (peak voltage dee to ground). Cavities operate up to 850 kV per gap. The acceleration system often runs on harmonic of beam orbital frequency  $f_{rf} = hf_0$ . Multiple dees are sometimes phased by selecting a natural mode of the resonator structure (0-mode or  $\pi$ -mode) and sometimes by servoed phase shifters (as in 3 dee  $h = 1$  systems at MSU, Texas A&M, Catania, & AGOR).

Without flutter,

$$\begin{aligned} \langle B \rangle_{isoc} &= 2\pi f_0 \frac{m_0}{q} \gamma(r) \\ &= 2\pi f_0 \frac{m}{q} \left[ 1 - \left( \frac{2\pi f_0 r}{c} \right)^2 \right]^{-1/2} \end{aligned} \quad (5)$$

$$k_{isoc} = \frac{r}{B} \frac{dB}{dr} = \gamma^2 - 1 \Rightarrow \nu_r = \gamma \quad (6)$$

As scalloping increases, the radial tune  $\nu_r$  rises faster than  $\gamma$ , hitting resonances. The  $N/2$  stopband limits maximum  $\gamma$  (e.g.,  $\gamma_{max} \approx 1.2 \Rightarrow 200$  MeV/A for  $N = 3$ ).

Axial focusing tune  $\nu_z \approx [-k + F(1 + 2\tan^2 \alpha)]^{1/2}$ . Normally one picks  $\alpha(r)$  to increase with  $r$  to give  $\nu_z \approx 0.2$  to avoid resonances. If sectors are radial ( $\alpha = 0$ ),  $\nu_z$  decreases with  $r$  and resonances limit usable energy band.

Due to large energy gain per turn and relatively rapid change of  $\nu$  values with energy, cyclotrons frequently pass through *essential* betatron resonances at  $\nu = N/3, N/4$ , etc. and *imperfection* resonances at 1/1, 2/2. Imperfection resonances are often intentionally used to steer the beam in the central region and near extraction.

### Space Charge Limits

Axial limit Assume fully overlapping turns ("current sheet" approximation),

$$I_{lim} = \epsilon_0 (2\pi f_0) A \nu_{z,min}^2 D_f V \quad (7)$$

where  $\epsilon_0$  is the permittivity of free space,  $A$  is the full beam height,  $\nu_{z,min}$  is the axial tune at the radius of weakest focusing,  $D_f$  is the fraction of the machine circumference occupied by the beam, and  $V$  the peak accelerating voltage per turn.  $I_{lim} \approx 10$  mA is typical.

Longitudinal limit In an isochronous device the orbit period is independent of the energy so the longitudinal length of bunch is unchanged but space charge force increases the energy spread of particles in a given turn and is the principal limit for single-turn-extraction (see below). The linear component of space charge force can be compensated by moving beam slightly to side of voltage wave. Numerical calculations are required to estimate residual nonlinear component. A scaling law shows that turn separation  $\propto V^3$ , so  $V$  is critical. "Flat-topping" voltage by adding higher harmonic gaps is sometimes used.

**Beam extraction**  $H^-$  cyclotrons use stripping foils. Positive ion cyclotrons use an electrostatic deflector with thin ( $\sim 0.1$ -0.3 mm) septum followed by magnetic channel(s).

Precessional extraction introduces coherent radial oscillation (amplitude  $a$ ) prior to extraction to add precessional component ( $2\pi a |\nu_r - 1|$ ) to the acceleration radius-gain-per-turn [ $dr/dn = r(\gamma/(1+\gamma))(qV/E)(\nu_r)^{-2}$  where  $qV$  is the energy-gain-per-turn]. Precession usually induces a  $\nu_r = 1$  transition at the edge region using a field bump (i.e. magnetic azimuthal first harmonic) of  $\sim 10^{-5}$  relative to main field. A radius gain of 3-5 mm is typical (limited by axial instability at the nearby  $\nu_r = 2\nu_z$  coupling resonance).

Single-turn extraction uses high accelerating voltage  $V$  and narrow rf phase interval at voltage peak to give turn spacing larger than turn width so that the deflector septum can be placed at a low density point between last two turns (extraction efficiency of 99.97% at 590 MeV at PSI). This requires highly stabilized magnet and rf (amplitude

and frequency). Flat-topping the rf waveform by adding a third harmonic is used at PSI to broaden the usable phase interval. Extraction current is limited by longitudinal space charge spreading of the turn structure. Very high  $V$  is required at high currents ( $\sim 3.4$  MV/turn at PSI for a 2.2-mA beam - until recently at 1.3 MW currently the highest power beam from any accelerator, cw or pulsed).

Multi-turn extraction is the default situation when single-turn requirements are not met. Particles on either side of the minimum-turn-number phase make an additional turn and so on to the edge of occupied phase interval. Particles at turn number steps line up on the septum and a fraction of (septum thickness)/(radius-gain-per-turn) is lost. Overall multi-turn extraction efficiency ranges from 50% (typical) to 93% (PSI) to 97% (numerical optimization).

## References

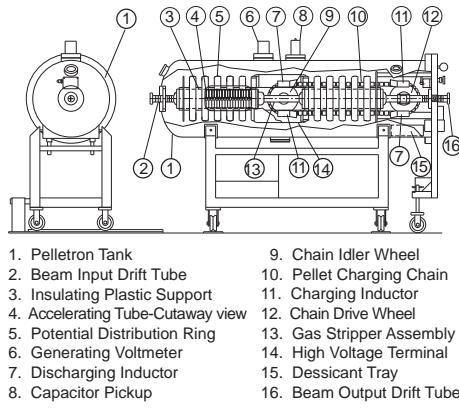
- [1] Lawrence, Edlefsen, Science LXXII (1930) 376
- [2] Proc. Int. Conf. on Cyclotrons and Their Applications, <http://jacow.org/>
- [3] M.K. Craddock, K. Symon, RAST 1, 65 (2008)
- [4] L.H. Thomas, PR 54 (1938) 580
- [5] M.E. Rickey, W.R. Smythe, NIM 18-19, 66 (1962)

### 1.6.5 Electrostatic Accelerator

*J. Ferry, NEC*

An electrostatic accelerator is a single potential drop system in which the high voltage potential is generated by the mechanical transfer of charge from ground to the high voltage terminal. Examples: fabric belt charged Van de Graaff accelerators built by High Voltage Engineering Europa B.V., the Netherlands, by Nissin-High Voltage Co., Ltd., Japan, and by Vivirad High Voltage, USA, and pellet chain charged Pelletron® built by National Electrostatics Corp (Fig.1).

Most electrostatic accelerators built today are dual acceleration, tandem types. Negative ions are injected and accelerated by a positive dc voltage on the high voltage terminal. At MeV energies, the ions are directed through a turbo-molecular pumped [1], i.e. compressed, tube of argon or nitrogen gas (microns Hg), which strips two or more electrons away changing them to positive ions. Carbon foils of a few  $\mu\text{gm}/\text{cm}^2$  can also be used, however important heavy ions such as  $\text{C}^{+3}$  at 2.6 MeV cause the foils to rapidly thicken and break. The positive ions are accelerated away



11-97 835SA46

Figure 1: 1 MV Tandem Pelletron accelerator.

Table 1: Tandem Pelletron® basic specifications.  $V$  is terminal voltage,  $D$  and  $L$  are pressure tank diameter and length,  $\hat{I}$  is maximum charging current. (\*Folded tandem with 180° magnetic deflection in terminal.)

| $V(\text{MV})$ | Orient. | $D(\text{m})$ | $L(\text{m})$ | $\hat{I}(\mu\text{A})$ |
|----------------|---------|---------------|---------------|------------------------|
| 1              | Horiz.  | .61           | 2.18          | 50                     |
| 3              | Horiz.  | 1.22          | 5.11          | 300                    |
| 5              | Horiz.  | 2.13          | 8.43          | 1000                   |
| 10             | Vert.   | 4.27          | 18.3          | 300                    |
| 15             | Vert.   | 5.49          | 26.3          | 300                    |
| 20             | Vert.*  | 8.23          | 23.77         | 900                    |
| 25             | Vert.*  | 10.06         | 29.87         | 900                    |

from the high voltage terminal to ground gaining additional energy. Tab.1 shows tank size versus terminal voltage for tandem pelletrons.

The FWHM of the terminal voltage variation of a Pelletron is  $\sim 500$  V without special ripple reduction circuitry. Thus, for a proton beam the FWHM energy spread is 500 eV. For a  $\text{C}^{+4}$  beam it is 2000 eV.

Final emittance depends on ion mass because of multiple scattering in the terminal stripper. For light ions ( $\text{p}'$ s and  $\alpha$ 's), final emittance is  $\sim 3\pi \text{ mm-mrad (MeV)}^{1/2}$ . For heavy ions like gold, the final emittance will be limited by the acceptance of the high energy acceleration tube to about  $20\pi \text{ mm-mrad (MeV)}^{1/2}$ .

**Acceleration tube** All electrostatic accelerators must have a highly evacuated tube for the ions to traverse during acceleration. This tube is made up of dense, low outgassing, insulating rings sealed

to metal electrodes. The voltage between the electrodes is established by resistors or corona points draining current from the high voltage terminal. Glass insulating rings sealed with polyvinylacetate, together with aluminum, titanium or stainless steel electrodes are successfully used. NEC tube uses high density alumina ceramic insulating rings sealed with aluminum metal to titanium metal electrodes. For most acceleration tubes the electric fields along the tubes are inclined or spiraled so that low energy secondary ions and electrons produced on the electrodes are deflected into nearby electrodes and cannot gain high energies [2, 3]. Some tubes also use low magnet fields to sweep low energy secondaries and electrons out of the beam path [4]. To remove ions and electrons produced on the electrodes, one can also vary the electric field along the beam path to produce cylindrically symmetric focusing fields that deflect low energy particles into adjacent electrodes [5, 6]. It is generally found that a comfortable reliable gradient of about 1.6 MV/m can be achieved with a modern acceleration tube. See also Secs.6.14, 7.2.9.

**Support column** Below 6 MV terminal voltage, one can use acrylic plastic plates as the insulating mechanical support column for components inside the pressure tank. Aluminum hoops are positioned along the insulating plates and are connected to the acceleration tube resistor divider [7]. This construction is suitable for columns up to about 29 ft. length. Above 6 MV, columns with alumina ceramic insulators can be used. In this case ceramic cylinders are bonded with aluminum to titanium electrodes to form posts about 18 in. long. The posts can support 1 MV when immersed in pressurized SF<sub>6</sub> gas.

**Charging system** Many fabric belts are still in use, although no new machines do so. Since the fabric belt is an insulator, charge transfers to and from the belt must be accomplished via corona discharge or physical rubbing. Such methods, although effective, are difficult to control precisely, leading to large terminal voltage ripple and can limit useful belt life.

Another method uses a chain made up from steel cylindrical pellets linked by stainless steel pins to nylon insulators. A chain carries typically  $> 150 \mu\text{A}$  of current [8]. The chains are charged and discharged by an induction scheme. Each chain is surrounded by a long, closely spaced inductor electrode where the chain leaves or arrives

at a wheel. The inductor electrodes are biased up to  $-60 \text{ kV}$  for chain runs to and  $+60 \text{ kV}$  for chain runs from a positive polarity high voltage terminal. As the grounded chain, pressing firmly against a wheel, moves into the inductor electrode charge flows smoothly onto it. Then while the chain is still within the inducing field, contact with ground is broken as the wheel turns. The induced charge is trapped on the chain and carried to the terminal or ground where the chain enters the discharging inductor electrode. Contact with the wheel occurs as the charge on the pellet is bound by the inductor to prevent sparks or corona discharges from the chain.

## References

- [1] F. Terrasi et al, NIM B259 (2007) 15
- [2] W.D. Allen, Nat. Inst. Res. Nucl. Sci. NIRL/R/21 (1962)
- [3] R.J. Van de Graaff, P.H. Rose, A.B. Wittkower, Nature 195 (1962) 1292
- [4] F.A. Howe, IEEE Trans. Nucl. Sci. N.S.-14, No.3 (1967) 122
- [5] M.L. Sundquist, R.D. Rathmell, J.E. Raatz, NIM A287 (1990) 87
- [6] J.E. Raatz et al, NIM A244 (1986) 104
- [7] J.A. Ferry, NIM A328 (1993) 28
- [8] G.A. Norton et al, NIM B37/38 (1989) 403

## 1.6.6 FFAG Accelerator

*M.K. Craddock, U. British Columbia & TRIUMF*

Following the discovery of alternating gradient (AG) focusing in 1952, Fixed-Field Alternating-Gradient accelerators (FFAGs) were proposed independently by Ohkawa in Japan, Kolomensky in the USSR and Symon and Kerst in the US [1]. With fixed magnetic fields, modulated rf, and pulsed beams, FFAGs operate just like synchrocyclotrons, but are capable of reaching tens of GeV. The innovations were to break the magnet into radial or spiral sectors to provide strong focusing, and (usually) to remove the central region — the same steps that convert a classic Lawrence cyclotron into a separated-sector ring cyclotron. The FFAG is the most general type of fixed-field accelerator (*i.e.* cyclotron [2], Sec.1.6.4).

Fixed magnetic fields lead to spiral orbits, so an FFAG's vacuum chamber, magnets and rf cavities tend to be larger and more costly than a synchrotron's. On the other hand, its beam intensity can be much higher, as the radial and momentum

Table 1: The cyclotron family

| Magnetic field variation $B(\theta)$ | Fixed frequency (CW beam) | Freq. modulated (Pulsed beam) |
|--------------------------------------|---------------------------|-------------------------------|
| Uniform                              | Classical                 | Synchro-                      |
| Periodic                             | Isochronous               | FFAG                          |

acceptances are larger, and the repetition rate, set purely by rf considerations, can be several kHz.

The most intensive studies were carried out by Symon, Kerst and others at MURA (the Mid-western Universities Research Association) in Wisconsin in the 1950s and 60s, and culminated in the construction and successful testing of electron models of radial-sector and spiral-sector designs [3]. But proposals for proton FFAGs were not funded at that time, nor were those for 1.5 GeV spallation neutron sources in the 1980s.

Recently, with improvements in magnet and rf technology, FFAGs have become the focus of renewed attention. Ten machines have been built and a muon cooling ring is under construction. In addition, ~15 designs are under study for the acceleration of protons, heavy ions, electrons and muons, with applications as diverse as treating cancer, irradiating materials, driving subcritical reactors, boosting high-energy proton intensity, and producing neutrinos. Moreover, it has become apparent that FFAG designs need not be restricted to the “scaling” approach explored in the 1950s. Dropping this restriction has revealed a range of interesting new design possibilities, which have been explored in a series of FFAG Workshops [4]. For recent reviews see [5, 6].

**Scaling FFAGs** Resonance crossing was a big worry in the early days of AG focusing, because of the low energy-gain/turn. The *scaling* principle was therefore adopted, whereby the orbit shape, optics and tunes are kept the same at all energies. To first order the tunes are given by

$$\nu_r^2 \approx 1 + k \quad (1)$$

$$\nu_z^2 \approx -k + F(1 + 2 \tan^2 \alpha) \quad (2)$$

where

- average field index  $k(r) \equiv r(d\bar{B}/dr)/\bar{B}$
- average field at radius  $r$  is  $\bar{B} \equiv \langle B(\theta) \rangle$
- magnetic flutter  $F \equiv \langle (B(\theta)/\bar{B} - 1)^2 \rangle$
- sector spiral angle  $\alpha \equiv r(d\theta/dr)$ .

Clearly, constant  $\nu_r$  requires  $k = \text{constant}$ , implying a magnetic field profile  $\bar{B} = B_0(r/r_0)^k$  and a momentum profile  $p = p_0(r/r_0)^{k+1}$ . As a large  $k$  is usually chosen to minimize the radial aperture, constant  $\nu_z$  requires a constant, high, value

of  $F(1 + 2 \tan^2 \alpha)$ . MURA’s recipe was to keep the flutter  $F(r) = \text{constant}$  by using constant profile  $B(\theta)/\bar{B}$ , and, depending on the sector type:

- if spiral:  $\alpha$  constant, sector axis  $r = r_0 e^{\theta \cot \alpha}$
- if radial: boost  $F$  by specifying  $B_D = -B_F$ .

Of course, reverse fields raise the average radius. The “circumference factor”  $R/\rho \geq 4.5$  if there are no straights [1], but smaller with them (1.8 for the KEK 150 MeV ring).

**Scaling FFAGs operating or under construction** Recent years have seen the construction and successful operation of the first-ever FFAGs for protons by Mori’s group at KEK [7, 8] and several more following scaling principles (Tab.2). Except for ERIT and NHV (FDF), the radial-sector designs employ DFD triplet magnets.

Mori introduced important innovations in both magnet and rf design. The DFD triplets are built and powered as single units, without a steel return yoke, forcing the return flux through the D and automatically providing reverse field. The open structure also facilitates injection and extraction. The rf innovation (avoiding the cumbersome rotary capacitors on synchrocyclotrons) is the use of FINEMET metallic alloy tuners, which offer (a) rf modulation (with a 1.5–4.6 MHz sweep) at 250 Hz or more, and so high beam-pulse rep rates; (b) high permeability, and so short cavities with high effective fields; (c) low  $Q$  ( $\approx 1$ ), allowing broadband operation.

Table 2: Scaling FFAGs operating or being built

|            | E<br>(MeV) | Ion Cells | Spiral angle | Radius<br>(m) | First beam |
|------------|------------|-----------|--------------|---------------|------------|
| KEK-PoP    | 1          | p         | 8            | 0°            | 0.8-1.1    |
| KEK        | 150        | p         | 12           | 0°            | 4.5-5.2    |
| KURRI      | 2.5        | p         | 8            | 40°           | 0.6-1.0    |
| ADSR rings | 20         | p         | 8            | 0°            | 1.3-1.9    |
|            | 150        | p         | 12           | 0°            | 4.5-5.1    |
| ERIT       | 11         | p         | 8            | 0°            | 2.35       |
| PRISM test | 0.8        | α         | 6            | 0°            | 3.3        |
| PRISM      | 20         | μ         | 10           | 0°            | 6.5        |
| NHV        | 0.5        | e         | 6            | 30°           | 0.19-0.44  |
| NHV        | 10         | e         | 8            | 0°            | 0.43-0.71  |
|            |            |           |              |               | 2011       |

A similar 150-MeV FFAG has been built at the Kyoto University Research Reactor Institute (KURRI), together with injector and booster FFAGs. The world’s first tests of accelerator-driven sub-critical reactor (ADSR) operation were carried out there in 2009. An 11-MeV 70-mA

## Sec.1.6: GLOSSARY OF ACCELERATOR TYPES

FFAG proton storage ring has also been built at KURRI for boron neutron capture therapy using an internal Be target, the high current being maintained by ionization cooling (Sec.2.7.4).

FFAGs are also of interest for muons. PRISM (Phase-Rotated Intense Slow Muon source), a 68 MeV/c DFD storage ring, is under construction at RCNP Osaka for eventual installation at J-PARC (Sato *et al.* [9]). The muon bunches it collects will be rotated in phase space, reducing their momentum spread from  $\pm 30\%$  to  $\pm 3\%$  for ultra-sensitive studies of rare muon decays. Optics measurements have been carried out with  $\alpha$ -particles.

**Scaling FFAG studies** In addition, a number of different scaling FFAG designs are being studied [6], some of spiral- and some of radial-sector design. These range from a fist-sized prototype of a 10-MeV machine for electron irradiation, to medium-sized sources for 230-MeV proton and 400 MeV/u ion therapy (for which the high pulse repetition rates are clinically advantageous), to a chain of four FFAGs accelerating muons from 0.3 to 20 GeV for a neutrino factory (see also Sec.1.6.15) at J-PARC.

For the 3.6-12.6 GeV stage of the reference scheme of the International Design Study of a Neutrino Factory (IDS-NF) Mori's group proposes a 161-m radius ring with 225 FDF cells; with 1.8 GV per turn at a fixed frequency of 200 MHz, acceleration would occur over 6 turns within a stationary bucket (T. Planche *et al.* [10] p.4506). A major advantage of FFAGs over linacs – either single or recirculating – is that their large acceptances in  $r$  and  $p$  reduce the need for muon cooling or phase rotation. Moreover the accelerators themselves are significantly less costly.

Fixed-frequency acceleration between stationary buckets and by harmonic number jumping (HNJ) have also been considered. The former “serpentine” method (see below) is possible in scaling FFAGs if the orbital period minimum at  $\gamma = \sqrt{k+1}$  lies within the acceleration range, and has been successfully demonstrated by Yamakawa *et al.* [4] on the NHV 10-MeV electron FFAG.

**Quasi-scaling FFAGs** A novel technique for creating a scaling field with superconducting magnets has been proposed by Machida ([10] p.112) for the PAMELA cancer therapy FFAG. This uses nested 2-, 4-, 6- and 8-pole solenoids to approximate the  $r^k$  law over the beam aperture.

A first 12-triplet ring (radius 6.25 m) would provide 250-MeV protons, and a similar second ring (radius 9.3 m) 400-MeV/u C<sup>6+</sup> ions.

**Vertical scaling FFAGs** A more radical approach has been taken by Brooks [4, 11] who proposes a “vertical FFAG” where the beam follows a helical path at fixed radius in a ring of superconducting magnets, and shows that a scaling law  $B_z \propto e^{kz}$  produces skew-quadrupole focusing and constant tunes.

**Linear nonscaling (LNS) FFAGs** In a study of FFAG arcs for recirculating linacs in 1997, Mills and Johnstone pointed out that the rapid acceleration (<20 turns) essential for muons allows betatron resonances no time to damage beam quality. Scaling can therefore be abandoned, the tunes allowed to vary, and a wider variety of lattices explored. Moreover, using constant-gradient “linear” magnets greatly increases dynamic aperture and simplifies construction, while employing the strongest possible gradients minimizes the physical aperture. Johnstone *et al.* [12] applied this *nonscaling* approach to a complete FFAG ring, showing that it would be advantageous to use superconducting magnets in which the field strengths decrease outwards - *i.e.* positively bending Ds and reverse-bending Fs (Fig.1). The radial orbit spread would be reduced (allowing the use of smaller vacuum chambers and magnets), and the orbit length  $C(p)$  shortened and made to pass through a minimum instead of rising monotonically as  $p^{1/(k+1)}$  (Fig.2 (left)). The variation in orbit period is thereby reduced, allowing the use of high- $Q$  fixed-frequency rf.

$C(p)$ ’s parabolic variation and its parametric dependence can be derived by treating the F and D

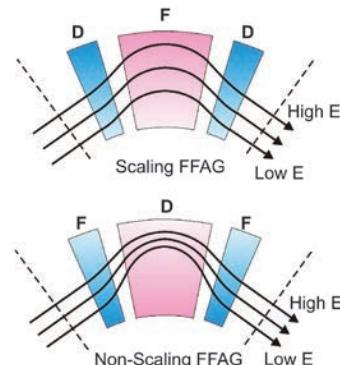


Figure 1: Scaling and nonscaling FFAG orbits.

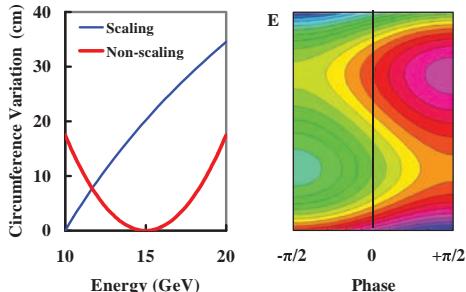


Figure 2: Linear nonscaling FFAGs: (left) circumference variation with energy; (right) acceleration path (yellow) in longitudinal phase space.

magnets as thin lenses of strength  $q_i$  (gradient  $\times$  length) [13]. For symmetric F0D0 or triplet cells, and assuming  $q_F = q_D = q$ ,

$$C(p) = C(p_m) + \frac{12\pi^2}{e^2 q^2 N L_{FD}} (p - p_m)^2 \quad (3)$$

where  $N$  is the number of cells, and  $L_{FD}$  is the (shorter) F-D spacing. The minimum is at  $p_m = (4p_c + eqL_{FD})/6$  where the  $p_c$  closed orbit is such that  $B_F = 0$ . The orbit radii  $r(p)$  show similar dependence, with distinct  $p_{\min}$ .

Lattices along these lines have been developed. The IDS-NF has adopted a LNS-FFAG for the final 12.6-25 GeV stage of muon acceleration (67 FDF cells, radius 111 m, decay loss 7.0% over 11.6 turns) [14].

With the orbit length varying by only 0.02%, first falling and then rising, Berg and Koscielniak [15] have shown that by exceeding a critical rf voltage an acceleration path can be created (Fig.2 (right)) that stays close to the voltage peak (crossing it three times), snaking between neighbouring buckets (rather than circulating inside them) just as in an imperfectly isochronous cyclotron. In order to demonstrate the novel features of such a design – particularly the serpentine acceleration outside buckets, and the crossing of many integer and half-integer resonances – a 10-20 MeV electron model (EMMA) has been built [16] and successfully commissioned at Daresbury. Initial experiments on these two features have shown that the beam behaves as expected [17], confirming the viability of the LNS approach.

LNS-FFAGs have also been considered for lower-energy applications with slower acceleration, where resonance crossing is of more concern. Trbojevic *et al.* ([5] p.115) have proposed

a LNS-FFAG complex of three concentric 48-cell rings to produce 250-MeV protons and 400-MeV/u  $C^{6+}$  for cancer therapy, and show that dangerous resonances can be avoided. They also propose ([10] p.124) a lightweight LNS-FFAG gantry, capable of accepting the whole extracted momentum range at fixed field. This would be composed of either superconducting or permanent magnet triplets, and weigh only  $\sim 1.5$  tons.

Ruggiero ([5] p.84) proposed a chain of 3 LNS-FFAGs in the AGS tunnel as a 19-MW, 12-GeV proton driver. Operation could be either pulsed (100 Hz) or c.w. (requiring harmonic number jumping).

**Nonlinear nonscaling (NLNS) FFAGs** For cancer therapy, the high rf voltage needed to cross resonances quickly would be too expensive. Johnstone's "tune-stabilized" design [18] keeps the betatron tunes constant by using NLNS magnets where the pole gap varies with radius and the sector edges are straight, but not all radial, providing extra edge focusing. There are separate 250-MeV proton and 400-MeV/u  $C^{6+}$  rings, each with 4 triplet cells. A similar 1000-MeV proton ring could be added to serve as an ADSR driver.

Rees ([5] p.74) also uses nonlinear field profiles but a more complicated dFDFd cell structure, to design a muon ring that is isochronous from 8 to 20 GeV – a muon cyclotron. The extra magnets provide an additional degree of freedom, rendering Eq.(2), derived for two-component cells, no longer applicable, so that spiral edges are not required. Méot *et al.* ([5] p.44) have carried out tracking studies without any attempt at fine tuning and find only small losses at a few resonances.

Using a similar lattice Rees has also designed a non-isochronous 4-MW 10-GeV proton driver ( $C = 624$  m) ([5] p.102). This would operate at 50 Hz, fed by a 3-GeV RCS.

**Insertions** Beam injection and extraction is often a design challenge for FFAGs, because of the regular cell structure. Insertions providing longer drift spaces can however be designed, as first shown for scaling FFAGs by Meads [19]. Machida ([10] p.558) has recently presented a design with 4 straights that are well-matched to the arcs over the full 30-250 MeV energy range. Lagrange and Mori ([10] p.4503) have proposed racetrack versions of the PRISM and ERIT storage rings, and have built a single straight scaling FFAG triplet for beam tests [4]. For NS-FFAGs Rees ([5] p.74) showed that four 9-cell straights

## Sec.1.6: GLOSSARY OF ACCELERATOR TYPES

could be inserted in his isochronous muon ring, matched acceptably to the main arcs at all momenta. Concentration of the rf allowed the circumference to be reduced from 1255 m to 903 m.

### References

- [1] K.R. Symon et al, PR 103 (1956) 1837
- [2] E.M. McMillan, in "Experimental Nuclear Physics", v.3, 639-786 (Wiley, 1959)
- [3] K.R. Symon, PAC'03, 452 (2003); F.T. Cole, Proc. Cyc'01, Suppl. (2001)
- [4] FFAG'11 Workshop;  
<http://www.cockcroft.ac.uk/events/ffag11>
- [5] C.H Prior (ed.) ICFA Beam Dynamics Newsletter 43, 19-133 (2007);  
[www-bd.fnal.gov/icfadb/Newsletter43.pdf](http://www-bd.fnal.gov/icfadb/Newsletter43.pdf)
- [6] M.K. Craddock, K.R. Symon, RAST, 1, 65 (2008).
- [7] M. Aiba et al, EPAC'00, 299
- [8] S. Machida et al, PAC'03, 3452
- [9] A. Sato et al, EPAC'06, 2508; EPAC'08, 3389
- [10] Proc. IPAC'10, <http://www.jacow.org>
- [11] S.J. Brooks, HB2010, 96,  
<http://hb2010.web.psi.ch/>
- [12] C. Johnstone, W. Wan, A. Garren, PAC'99, 3068
- [13] M.K. Craddock, FFAG2003; S. Koscielniak, M.K. Craddock, EPAC'04, 1138
- [14] J.S. Berg et al, PAC'12, 241
- [15] J.S. Berg, Proc. Snowmass 2001, T503 (2001); S. Koscielniak, C. Johnstone, ibid., T508 (2001)
- [16] R. Barlow et al, NIM A624, 1 (2010)
- [17] S. Machida et al, Nature Physics 8, 243 (2012) | doi:10.1038/nphys2179
- [18] C. Johnstone et al, PAC'11, 2116 (2011)
- [19] P.F. Meads Jr., PAC'83, 244

### 1.6.7 High Voltage Electrodynanic Accelerators

*M.R. Cleland, IBA Industrial, Inc.*

High voltage electrodynanic accelerators are also called potential-drop or direct-action systems. They increase the kinetic energies of ions and electrons by connecting particle sources to high voltage generators and accelerating the particles to a variety of targets at ground potential. In contrast to electrostatic accelerators (Sec.1.6.5) that transport charges mechanically from ground to the particle source, the high voltage power supplies in electrodynanic accelerators convert low voltage alternating current (ac) to high voltage direct current (dc) by means of cascaded rectifier circuits. Some of these systems transfer ac

power from the electric service to multiple rectifier stages through an array of high voltage capacitors, while others use an assembly of high voltage transformers [1, 2, 3]. Although there are several different designs of the high voltage power supplies, the particle sources and acceleration methods are usually similar.

**Particle acceleration** Charged particles are accelerated in highly evacuated tubes to minimize collision with residual gases. Particles are extracted from their sources (ions from plasmas, electrons from hot cathodes), focused, and accelerated by strong electric fields created by the high voltage potentials. The contours and strengths of the electric fields are determined by the shapes and spacings of the accelerating electrodes with intermediate potentials. The use of multiple overlapping electrodes prevents spark discharges in the acceleration tube, protects the insulating rings between the electrodes from scattered particles and permits the use of high potentials and strong electric fields.

**High voltage generation** The many applications can require potentials from as low as 70 kV to as high as 5 MV. For voltage ratings below 300 kV, conventional single-stage transformer-rectifier systems can be used, but for higher voltages, multiple-stage cascaded rectifier systems are needed to insulate the secondary dc circuits from the primary ac power source. A variety of such multiple-stage systems have been developed. The different methods for coupling ac power to all of the rectifier stages are illustrated in Fig.1 [2]. In Fig.1a, ac power is inductively coupled in series from each of the transformer-rectifier stages to the next stage. The dc outputs of each stage are connected in series to produce high voltage dc power. In Fig.1b, ac power is inductively coupled in parallel from a common primary winding to all of the secondary windings. In Fig.1c, ac power is capacitively coupled in series from each rectifier stage to the next. Transformers are not needed in the rectifier stages. In Fig.1d, ac power is capacitively coupled in parallel to the rectifier stages.

In series-coupled systems, the lower stages in the cascaded rectifier circuit must transmit ac power to the upper stages, so the dc voltage generated under load is reduced in the upper stages. In parallel-coupled systems, all stages receive ac power directly from the primary source and the voltage generated under load is the same in

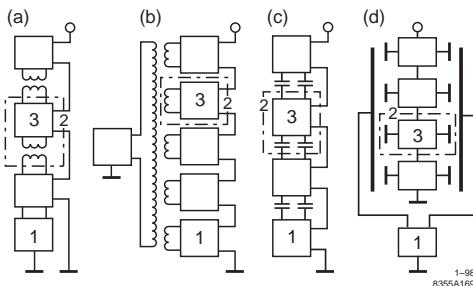


Figure 1: Cascade generators. 1—Power supply; 2—stage; 3—voltage rectifying and multiplying circuits.

all stages. Parallel coupling allows the use of many more rectifier stages, reduces the internal impedance of the high voltage generator and increases the amount of current and power that can be provided for particle acceleration.

### Capacitive Cascade generators

Series-coupled systems Cockcroft and Walton [4] used a high voltage generator with multiple rectifier stages capacitively coupled to a source of ac power. This type of system is commonly called a Cockcroft-Walton accelerator. Their series-coupled voltage multiplying rectifier circuit was proposed earlier by Greinacher [5]. During the 1930s and 1940s, the firm N.V. Philips, the Netherlands, produced many air-insulated high voltage generators and accelerators based on this concept with voltages up to 3.0 MV [6].

During the 1950s, Emile Haefely & Co Ltd, Switzerland, developed a symmetrical series-coupled cascade circuit with three columns of capacitors. The middle column reduced the ac ripple voltage at the high voltage terminal. They produced many accelerators using this type of generator for a variety of research applications such as electron microscopy, ion injection into higher energy rf accelerators, and separation of high energy particle beams. Potentials up to 4.0 MV have been obtained with compressed gas insulation [7, 8]. During the 1980s, the firm Nissin High Voltage Co Ltd, in Japan, simplified the symmetrical cascade circuit by omitting the middle column of capacitors for industrial applications that do not need low ac ripple voltage. Voltages up to 5.0 MV have been obtained with compressed gas insulation. Their largest accelerator of this type is rated for 30 mA of dc electron beam current or 150 kW of beam power [9, 10].

Parallel-coupled systems The parallel-coupled cascade circuit proposed by Schenkel [11] predated the Greinacher circuit, but it was not used for very high voltage generators because of the difficulty of making capacitors that could withstand the dc output voltage. During the 1950s and 1960s, Radiation Dynamics, USA, developed the parallel-coupled cascade circuit used in Dynamitron accelerators [12]-[15]. High voltage, high frequency ac power at about 100 kHz is generated by a triode-driven resonant system consisting of an iron-free transformer and a pair of semicylindrical electrodes which surround the rectifier assembly. Power is capacitively coupled from these electrodes to semicircular corona rings connected to the rectifier junctions. The rectifiers are connected in series to produce high voltage dc power. Insulation between components is provided with compressed SF<sub>6</sub> gas. Dynamitrons with voltages up to 5.0 MV have been produced with electron beam power ratings up to 300 kW. Two-stage tandem ion accelerators using this type of high voltage generator have been produced by Radiation Dynamics, Inc. [16, 17]. General Ionex Corporation also produced parallel-coupled tandem accelerators called Tandetrons for high energy ion implantation in silicon wafers [18]. Similar systems are now made by High Voltage Engineering Europa in the Netherlands.

### Inductive Cascade generators

Series-coupled systems In the 1950s and 1960s, High Voltage Engineering Corporation developed the Insulating Core Transformer (ICT). This is a three-phase, multi-stage rectifier cascade circuit using magnetic coupling to transfer low voltage, low frequency ac power from the primary windings at the low voltage end of the transformer to an array of high voltage secondary windings. The magnetic cores of the secondary windings are separated by thin sheets of solid insulating material. Rectifiers and filter capacitors convert ac power to dc at each stage. All of the dc circuits are connected in series to produce the high voltage at the output end of the assembly [19]-[21]. The largest ICT accelerators produce voltages up to 3.0 MV with electron beam power ratings up to 100 kW. ICTs are now made by Vivirad High Voltage in France and Wasik Associates in the USA. Sealed Air Corp. makes 0.5 MV ICTs for irradiating their heat-shrinkable food packaging films.

Parallel-coupled systems During the 1970s, the Budker Institute of Nuclear Physics developed

## Sec.1.6: GLOSSARY OF ACCELERATOR TYPES

several types of single-phase, multi-stage transformer systems. All of the secondary windings are coupled to an external coaxial primary winding which extends the full length of the high voltage assembly. The Elita is a resonant pulse transformer with a solenoidal high voltage secondary winding. The ELT is a low frequency system for generating high voltage dc power. It has insulated magnetic cores inside the multiple secondary windings. The ELV is a similar multi-stage, transformer-rectifier system without magnetic cores. ELV accelerators produce electron beam power ratings of 50 kW with dc potentials up to 2.0 MV and 400 kW of beam power at 0.4 MV [2, 22, 23]. A 2.5 MV dc system has been made by LBNL [24].

During the 1970s and 1980s, the Budker Institute, the Efremov Institute for Electrophysical Apparatus in St. Petersburg, and the Institute for High Temperatures in Moscow developed three-phase, transformer-rectifier high voltage systems. These use continuous long iron cores at ground potential with primary windings on each core. The high voltage secondary windings are insulated from the cores and primary windings. Rectifiers connected in three-phase bridge circuits convert ac to dc power. The Teus system uses a single bridge rectifier circuit with three high voltage windings and can produce 0.3 MV dc and 75 kW of electron beam power. The Aurora system uses multiple bridge rectifier circuits with lower voltage per stage and can produce 0.75 MV dc and 100 kW of beam power [2, 25, 26, 27]. More powerful systems have been developed by the Efremov Institute for applications that do not need voltages higher than 1.0 MV.

## References

- [1] W. Scharf, Particle Accelerators and Their Uses, Part 1, Accelerator Design, Part 2, Applications of Accelerators, Harwood Academic (1986)
- [2] E.A. Abramyan, Industrial Electron Accelerators and Applications, Hemisphere (1988)
- [3] M.R. Cleland, Chapter in Radiation Processing of Polymers, Hanser (1992)
- [4] J.D. Cockcroft, E.T.S. Walton, Proc. Royal Soc. London, Series A136 (1932) p.619
- [5] H. Greinacher, Zeit. fur Physik, Vol.4 (1921) 195
- [6] A. Bouwers, A. Kuntke, Zeit. fur Technische Physik, Vol.18 (1937) 209
- [7] W. Heilpern, Helvetica Phys. Acta, Vol.28 (1955)
- [8] G. Reinhold, K. Truempp, J. Bill, IEEE Trans. Nucl. Sci., NS-12, No.3 (1965) 288
- [9] K. Mizusawa et al, Radiation Phys. & Chem., Vol.31, Nos.1-3 (1988) 267
- [10] S. Uehara et al, Radiation Phys. & Chem., Vol.42, Nos.1-3 (1993) 515
- [11] M. Schenkel, Elektrotechnische Zeitschrift, Vol.40 (1919) 333
- [12] M.R. Cleland, Voltage Multiplication Apparatus, U.S. Patent No. 2,875,394 (1959)
- [13] M.R. Cleland, P. Farrell, IEEE Trans. Nucl. Sci., NS-12, No.3 (1965) 227
- [14] M.R. Cleland, C.C. Thompson, H.F. Malone, Radiation Phys. & Chem., Vol.9, Nos.4-6 (1977) 547
- [15] C.C. Thompson, M.R. Cleland, NIM B40 (1989) 1137
- [16] P.R. Hanley et al, IEEE Trans. Nucl. Sci., NS-16 (1969) 90
- [17] S.A. Cox, P.R. Hanley, IEEE Trans. Nucl. Sci., NS-18 (1971) 108
- [18] K.H. Purser et al, SPIE Vol.530, Soc. Photo-Optical Instr. Eng. (1985) 14
- [19] R.M. Emanuelson, Insulating Core Transformers, U.S. Patent No. 3,274,526 (1966)
- [20] R.J. Van de Graaff, High Voltage Electromagnetic Apparatus Having an Insulating Magnetic Core, U.S. Patent No. 3,289,066 (1966)
- [21] R. Emanuelson, R. Fernald, C. Schmidt, Radiation Phys. & Chem., Vol.14 (1979) 343
- [22] G.I. Budker et al, Charged Particle Accelerator, SU Patent No. 589698 (1973), U.S. Patent No. 4,016,499, GB Patent No. 1454485
- [23] R.A. Salimov et al, Radiation Phys. & Chem., Vol.57, Nos.3-6 (2000) 661
- [24] L.L. Reginato et al, AIP Conf. Proc. 392 (1997) 1305
- [25] B.I. Albertinsky, M.P. Svin'in, Cascade Generators, Atomizdat (1980)
- [26] M.P. Svinin, Radiation Phys. & Chem., Vol.18, Nos.5-6 (1981) 1353
- [27] V.N. Lisin et al, High Voltage Transformer-Rectifier Device, U.S. Patent No. 4,338,657 (1982)

## 1.6.8 Induction Linac

*R. Bangerter, LBNL*

Induction linacs (IL) are employed in applications that require combinations of beam current, beam energy, and pulse lengths that are not easily achieved using rf accelerators. The first large IL was the Astron Injector at LLNL [1]. Although induction acceleration had been used for

some time (e.g. the betatron, Sec.1.6.2), Ref.[1] is usually credited for the invention of IL. Since then more than 40 ILs have been built. For a comprehensive review with listings of major ILs and their characteristics, see [2].

Conventional linacs produce long trains of beam pulses at radio frequencies. The beam current is usually  $< 1 \text{ A}$ . In contrast, ILs produce beam pulses at much lower repetition rates, but much higher currents (e.g.  $>10 \text{ kA}$ ). The Astron Injector originally produced 350 A of electrons (3.7 MeV, pulse length 300 ns, burst repetition rate  $>1 \text{ kHz}$ ) to create a magnetic field in the Astron magnetic fusion device. Subsequently ILs to produce beams for electron ring accelerators (experimental collective accelerators) were built at Berkeley and Dubna [3, 4]. Flash radiography is another important application of ILs, e.g. FXR and DARHT in the US, AIRIX in France, and DRAGON-I in China [2]. The ATA, a 50-MeV IL, originally built for defense studies, currently holds the record for the highest voltage (kinetic energy). Other applications that have been suggested or implemented include FELs, microwave power generation, electron beam welding, food irradiation, pulsed neutron sources, treatment of materials, treatment of chemical and nuclear waste, and tunneling in rock [2]. These applications use electron or proton beams. Heavy ion inertial confinement fusion and high energy density physics are applications that require beams of heavier ions (Li to U). The old hardware from the ATA has been modified to build NDCX-II, a lithium accelerator designed for basic studies of hot dense matter [5].

**The IL concept** Fig.1 (upper) illustrates the basic concept of induction acceleration. A pulser (or modulator) provides the power to energize an induction module. The electric acceleration field is confined primarily to the axis of the accelerator by the conducting walls – the induction cavity – surrounding the induction core. The core is usually made of ferromagnetic (or ferrimagnetic) material although it could be simply air or vacuum. One may view the induction module as an electrical transformer. The beam is the secondary of this transformer. An IL consists of one or more (often many) induction modules placed in series. One can also think of an induction cavity as a shorted transmission line as shown in Fig.1 (lower). In any case, the accelerating voltage  $V$  is associated

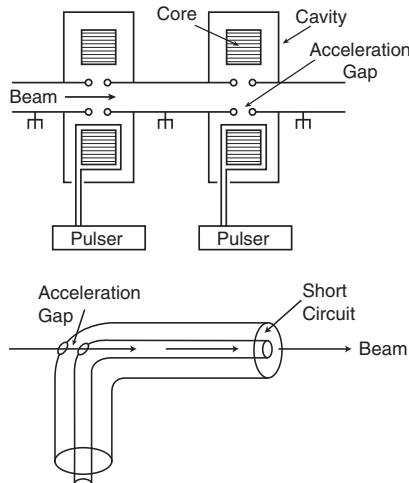


Figure 1: upper: An IL with two induction modules (stages). The beam is the secondary of a series of transformers. lower: Alternatively, one can think of an induction module as a shorted transmission line.

with a changing magnetic field by  $V = -d\Phi/dt$ , where  $\Phi = \int \vec{B} \cdot d\vec{A}$  is the magnetic flux in the core. The integral is over the cross-sectional area of the core. If one assumes that  $\vec{B}$  is uniform over the area, then  $V = -AdB/dt$ , or  $\int V dt = A \Delta B$ . This quantity is often referred to simply as the volt-seconds of the core, while  $\Delta B$  is referred to as the flux swing.

Another way to achieve induction acceleration is by changing the area occupied by the magnetic field rather than the field itself (line-type induction accelerator, in contrast to the core-type [3]). In the line-type machines, an induction cavity is basically a radial or axial transmission line. A ferromagnetic core is not used. The line-type cavities usually have very low impedance. Indeed, a principal function of the core in a core-type module is to provide a high impedance in parallel with the load (beam). In both induction accelerators and rf accelerators, the accelerating electric field is associated with a changing magnetic field. However, rf cavities are often driven at resonance and induction cavities are not. Also, induction cores are often large ( $\sim 1\text{m dia.}$ ). The large size, together with the use of ferromagnetic material, allows the induction accelerators to be used at longer pulse lengths than rf accelerators. For ILs, the pulse length lies in the range from tens of ns to tens of  $\mu\text{s}$ .

ILs are related to other devices, e.g., the induction voltage adder, where a solid conductor replaces the beam.

**Induction technology** The main components of an IL are the pulsers, the beam transport system, and, in the case of a core-type machine, the induction cores.

Depending on the voltage, pulse length, and other characteristics, the pulsers may be based on simple switched capacitors, pulse forming networks, or pulse forming lines. Switches for the pulsers include thyratrons, spark gaps, magnetic switches, and solid-state devices. To maximize  $\Delta B$ , a reset pulser normally magnetizes the core in one direction before it is pulsed in the opposite direction by the main pulser.

The beam transport system of an IL must be capable of carrying high current. One can estimate the maximum current that a transport channel can carry by setting the beam self force equal to the applied focusing force of the lattice. The actual current will be less than this maximum because of emittance. A detailed calculation is required in each individual case. For continuous solenoidal focusing the maximum current is approximately

$$I_S = 4 \times 10^5 (Z/A)\beta\gamma(Ba)^2 \text{ Amp.} \quad (1)$$

where  $Z$  and  $A$  are the charge and atomic mass numbers,  $B$  is in Tesla, and  $a$  is beam size in m. For a magnetic quadrupole transport system the maximum beam current is approximately

$$I_Q = 8 \times 10^5 (\beta\gamma)^2 (\eta Ba) \text{ Amp.} \quad (2)$$

where  $\eta$  is the effective occupancy factor (the fraction of the lattice occupied by quadrupoles). For electrostatic quadrupoles  $B$  is replaced by  $E/(\beta c)$  where  $E$  is the electrical field in V/m. Solenoids are usually preferred for light particles ( $e^-$ ,  $p$ ), but quadrupoles appear preferable for heavy ions, particularly at the high velocity end of the machine. There has been research on neutralized and collective focusing systems [6].

Magnetic materials used for induction cores include steel tape, nickel-iron tape, ferrite, and a class of amorphous metallic glasses such as Metglas®, originally produced by Allied Chemical, but now produced by Metglas® Inc., a subsidiary of Hitachi Metals. Core losses are an important consideration, particularly for lower beam currents. For core materials other than ferrite, the core must be laminated to minimize eddy-current

losses. Lamination is usually achieved by winding the cores from thin tape (thickness  $\leq 50 \mu\text{m}$ ). For example, typical losses in iron-based metallic glass are  $\approx 800 \text{ J/m}^3$  at a pulse length of  $\Delta t \approx 1 \mu\text{s}$  and a flux swing of 2.5 T. The losses scale approx. as  $\Delta B^2/\Delta t$ . (Because of hysteresis, the losses at 2.5 T do not drop appreciably below 100 J/m<sup>3</sup> even at very long pulse lengths.) A newer class of nanocrystalline materials such as Hitachi's Finemet and Vacuumschmelze's Vitroperm has losses that are typically several times lower than those of the older metallic glasses [7].

**Research topics** Many ILs have now been built, but IL technology is still not as highly developed as rf accelerator technology. For example, although DARHT has now achieved its design goals, during construction there were some unanticipated problems involving the behavior of ferromagnetic materials and the distribution of voltages within the induction cavities. This example points to the importance of continuing development of low-loss, low-cost ferromagnetic materials having predictable, consistent properties. In this regard, the nanocrystalline materials mentioned above appear to be promising. Also, since ILs often have relatively large apertures to carry high beam current, the development of large, inexpensive, highly reliable, high-gradient insulators is also an important research topic. Moreover, some IL applications such as high energy density physics and inertial fusion will likely require the development of novel beam transport systems such as compact, multi-beam arrays of superconducting quadrupoles. Research on this topic is in its infancy [8]. Finally, the development of advanced pulsers, e.g., those using solid state switches can be expected to have a profound influence on IL design and applications. For example, in one experiment [9], an uncooled induction core was run for a few weeks at a repetition rate of approximately 100 Hz. The pulser used capacitive storage and an array of field effect transistors. The measured efficiency into a resistive load approached 50%. With continued development, ILs may become competitive with rf accelerators for high average power applications.

## References

- [1] N.C. Christofilos et al, RSI 35 (1964) 886
- [2] Induction Accelerators, K. Takayama and R.J. Briggs (Eds.), Springer: NY, (2010)
- [3] J.E. Leiss, PAC 79, p.3870

- [4] A. Faltens, D. Keefe, Proc. Linac Conf., LANL Report, LA-9234-C (1981) p.205
- [5] A. Friedman et al, Physics of Plasmas 17, 056704 (2010)
- [6] S. Humphries, Jr. et al, PAC 81, p.3410
- [7] A.W. Molvik et al, LINAC98, p.320
- [8] G. Sabbi et al, NIM A 544 (2005) 285
- [9] W. Barletta et al, LBL-35960 (1994)

### 1.6.9 Industrial Accelerators

*R. Hellborg, Lund U.*

The years around 1930 can be taken as the starting point of the accelerator era when people conducted development work along different principles. In 1937 the first accelerator for applied use was constructed. This accelerator could, in over the energy range 0.5 – 1.2 MeV and electron beam up to 3 mA, produce x-rays for cancer treatment up to 0.01 C/kg<sub>air</sub> per mA beam current. Today accelerators are applied in very diverse fields. Of the more than 20 000 accelerators in operation around the world only a few hundred are used in (applied) physics research. One third is used in medical applications, two-thirds are used for industrial applications.

Electron beam processing is utilized by many major industries, including the plastics, automotive, rubber goods, petrochemical, wire and cable, electrical-insulation, textile, semiconductor, packaging and pollution control industry. In food industry electron beams are used for sterilization. Ion beam accelerators are used for a broad variety of industrial applications like micro-machining and for national security applications, which include x-ray inspection of cargo containers and stewardship of civil and military nuclear materials.

Tables 1, 2, 3, and 4 present some examples of industrial and medical applications of accelerators. The applications are given in column 1. In column 2 a suggested technique/method is given. In column 3 the necessary accelerator parameters are to be found. In column 4 suggestions for more readings are given.

The different types of accelerators used in industrial applications are:

1. The *cascade accelerator* (also called high voltage accelerator) in which the high-voltage unit consists of a multiplying rectifier-condenser system (first used by Cockcroft and Walton), see Sec.1.6.7.

2. In the *electrostatic accelerator* a charging system conveys the charge from ground to the insulated high voltage terminal (originally by an insulating belt, today mostly by a chain of metal cylinders connected by insulating links), see Sec.1.6.5.
3. In a *linear accelerator* (or linac) the beam travels through a series of hollow tubular electrodes. Along the Widerøe principle the electrodes are connected alternatively to opposite poles of an rf-supply. Along the Alvarez principle the structure consists of a set of resonator tubes having an rf-voltage of the same phase applied to them. Inside each resonator tube, a potential distribution exists, see Secs.1.6.11, 1.6.12.
4. The *RFQ accelerator* has a symmetry corresponding to that of an electrostatic quadrupole. It combines the action of focusing, bunching and acceleration, see Sec.1.6.18.
5. In the *cyclotron* the beam is bent into a circular path by a magnetic field and the particles orbit inside two semicircular metal chambers. These so called “dees” are connected to a source of alternating voltage. In the gap between the dees, the particles feel an accelerating voltage and gain energy twice during each circle, see Sec.1.6.4.
6. In the *synchrocyclotron* the frequency of the applied rf-field between the two dees is slowly decreased as the particle energy increases to compensate for the relativistic ion mass. Only one bunch at a time can be sent through this accelerator. (An alternative to a synchrocyclotron is to modify the cyclotron and divide the magnetic field into sectors of alternating high and low fields. This is called a *sector focusing* or *azimuthally varying field cyclotron*.) See Sec.1.6.4.
7. In the *synchrotron* the massive magnet (in cyclotrons and synchrocyclotrons) is replaced by a ring of bending magnets. The rf-field and the magnetic field strength are varied to keep the orbital radius of the beam constant. Only one bunch at a time is sent through the accelerator, see Sec.1.6.21.

Industrial and medical applications of accelerators will increase in the future and new applications can be foreseen [1]. New isotopes will be developed for the PET-technique. The isotope

## Sec.1.6: GLOSSARY OF ACCELERATOR TYPES

Table 1: Material processing by a high energy Ion Beam.

| Industrial/medical applications  | Suggested technique/ method to be used  | Accelerator parameters:<br>type <sup>a</sup> /particle /current /energy   | To learn more, see e.g.               |
|--|---|---|---------------------------------------|
| Fabrication of semiconductor devices and materials   | Ion implantation of dopants and other ions  | Casc, el stat, linac (rare cases RFQ) /several different /a few $\mu\text{A}$ to 100 mA /hundreds of keV (or even less) to MeV        | ch.1 in [3]                           |
| Filter and permeable membrane production   | Irradiation of thin films with heavy ions, followed by chemical etching of the films  | Casc, el stat /heavy ions /tens of $\mu\text{A}$ /10–100 MeV  | ch.17 in [2]                          |
| Material (nanoscale) engineering in fields like biomedical tissue engineering, quantum devices, optical and magnetic information storage technology etc. | Irradiation by MeV ions   | Casc, el stat /p, He, light and heavy ions up to Bi /1 fA – 100 mA /0.1 keV – 20 MeV  | ch.25 in [4], several chapters in [2] |
| Treatment of cancer tumors by radiation  | Irradiation with high doses of ions   | Cycl, synchrotr /p, $^{12}\text{C}$ / e.g. $^{12}\text{C}$ : 140–400 MeV $\text{u}^{-1}$  | [5]                                   |
| Mutogenesis of seeds for plant breeding  | Irradiation by keV – MeV ions   | Casc, el stat /heavy ions /e.g. $\text{N}^{2+}:10^{18}$ ions $\text{cm}^{-2}$ /e.g. $\text{N}^{2+}:30$ keV                            | [6]                                   |
| Clinical use of radio nuclides for therapy   | Radio nuclides produced by nuclear reactions and placed close to the tumor, either inside the body “brachytherapy” or close to the body “teletherapy” | Mainly cycl but even linacs are used/ mostly light ions like p, d, $^{3}\text{He}$ , $^{4}\text{He}$ /0.1 – 10 mA /mainly 10 – 20 MeV | ch.4 in [3], ch.19 in [4], [7]        |

<sup>a</sup> The abbreviations used for different types of accelerators are: casc = cascade accelerator; el stat = electrostatic accelerator; linac = linear accelerator; cycl = cyclotron (and synchro-cyclotron); synchrotr = synchrotron; RFQ = Radio Frequent Quadrupole accelerator. For detailed descriptions of different types of accelerators see sections in this handbook and [4].

$^{99m}\text{Tc}$  will be produced by accelerators. The need for  $^{99}\text{Mo}$  from nuclear reactors will therefore be reduced. A number of hadron (p and  $^{12}\text{C}$ ) therapy facilities are under construction. Hadron irradiation offers better dose distribution than conventional photon and electron beams do and for relatively radio-resistant tumors, treatment with heavier hadrons ( $^{12}\text{C}$  and  $^{20}\text{Ne}$ ) offers great potential benefit. An excellent depth dose can be reached

with energy- and intensity-modulated beams using 3-D scanning, Sec.1.6.14. To get quantitative information about sample composition and surface structure of nanometer technology products, a demand for MEIS (Medium Energy Ion Scattering) facilities can be foreseen. MEIS is performed at projectile energies of 100–300 keV. It involves very sophisticated/complicated detection equipment, see Ch.7 in [2].

Table 2: Material characterization by a high energy Ion Beam.

| Industrial/medical applications   | Suggested technique/ method to be used  | Accelerator type <sup>a</sup>  | parameters: /particle /current /energy | To learn more, see e.g. |
|---|---|--|--|-------------------------|
| Dating extremely small samples (down to 1 $\mu\text{g}$ or even less) for geological, cosmological and archeological purposes                             | Accelerator Mass Spectrometry (AMS)   | tandem casc, el stat/ $^{10}\text{Be}$ , $^{14}\text{C}$ , $^{26}\text{Al}$ , $^{36}\text{S}$ , $^{41}\text{Ca}$ etc./extremely low/tens of MeV or higher  | [8]                                    |                         |
| Quantifying extremely low concentrations of traces in small samples (down to 1 $\mu\text{g}$ or even less) for e.g. biomedical and environmental purposes | Accelerator Mass Spectrometry (AMS)   | tandem casc, el stat $^{10}\text{Be}$ , $^{14}\text{C}$ , $^{26}\text{Al}$ , $^{36}\text{S}$ , $^{41}\text{Ca}$ etc. /extremely low /tens of MeV or higher | [8]                                    |                         |
| Industrial use of radio nuclides, mainly for imaging and for tracking physical or biological processes in for example plants or animals                   | Radio nuclide production by suitable nuclear reactions                                | Mainly cycl but even linacs are used /mostly light ions like p, d, $^3\text{He}$ , $^4\text{He}$ /0.1 – 10 mA /mainly 10 – 20 MeV                          | ch.4 in [3], [7]                       |                         |
| Material analysis with bulk or depth sensitive nuclear methods  | Ion Beam Analysis (IBA) including methods like RBS, PIXE, PIGE, ERDA, CPA etc.        | single or tandem casc, el stat /light ions $/ \mu\text{A}$ /a few MeV  | ch.5 in [3]                            |                         |
| Pollution control (e.g. welding dust, combustion emissions, transport)  | Analyze with the PIXE-technique   | casc, el stat /protons $/ 1 - 10 \mu\text{A} / 2 - 3 \text{ MeV}$  | [9]                                    |                         |
| Geophysical exploration in the petroleum industry, moisture content measurements in timber and construction industry                                      | Detection of gamma-rays emitted from a neutron initiated nuclear reaction             | Casc or sealed tube generator /neutrons produced in a target after acceleration of p or d  | ch.6 in [3]                            |                         |
| Safeguard inspections for explosives, chemical weapons, fissionable material etc.   | Detection of neutrons or gamma-rays emitted from a neutron initiated nuclear reaction | Casc /neutrons produced in a target after acceleration of p or d to a few hundred keV  | ch.22 in [4], ch.6 in [3]              |                         |
| Biomedical use of radio nuclides, mainly for imaging and radiotracers to track physical or biological processes   | Radio-nuclide production by suitable nuclear reactions                                | Mainly cycl but even linacs are used /mostly light ions like p, d, $^3\text{He}$ , $^4\text{He}$ /0.1 – 10 mA /mainly 10 – 20 MeV                          | ch.19 in [4], [7]                      |                         |

<sup>a</sup>For the abbreviations used for different types of accelerators, see footnote in Tab.1.

## Sec.1.6: GLOSSARY OF ACCELERATOR TYPES

Table 3: Material processing by a high energy Electron Beam.

| Industrial/medical applications   | Suggested technique/ method to be used  | Accelerator parameters:<br>type <sup>a</sup> /particle /current /energy | To learn more, see e.g. |
|---|---|---|-------------------------|
| Sterilization, polymerization, disinfection, vulcanization and several other similar industrial processes | Irradiation by an electron beam with doses up to 100 kGy  | casc, el stat /electrons /10 $\mu$ A – 100 mA or more /up to a few MeV  | ch.28 in [4]            |
| Material processing like welding, cutting, drilling, heat treatment and melting                           | Electron beam irradiation   | casc, el stat /electrons /up to 1 A /up to 200 keV (in rare cases MeV)  | ch.2 in [3], [10]       |
| Treatment of cancer tumors by an external beam  | Irradiation with high doses of electrons or by x-rays obtained after the electron beam has collided with a target | linac /electrons /doses up to 2 Gy per treatment /10 – 20 MeV           | [11]                    |

<sup>a</sup>For the abbreviations used for different types of accelerators, see footnote in Tab.1.

Table 4: Material characterization by a high energy Electron Beam.

| Industrial/medical applications   | Suggested technique/ method to be used                   | Accelerator parameters:<br>type <sup>a</sup> /particle /current /energy   | To learn more, see e.g. |
|---|--|---|-------------------------|
| Production, quality control and control of regulatory requirements  | Spectroscopic and imaging technique by x-ray irradiation | Synchrotron producing x-rays of 1 eV – 100 keV (far infrared to hard x-ray)   | ch.8 in [3], [12]       |
| Computed tomography (CT) or 3-D scanning of industrial products. Metrology of internal structures of complex parts or assemblies. Non destructive testing and inspection of e.g. air cargos, hazardous waste etc. | x-ray or neutron irradiation                             | linac /electrons /absorbed dose up to tens of Gray/up to 10 MeV (quite low x-ray energy for micro and nanotomography) | ch.7 in [3], [13]       |

<sup>a</sup>For the abbreviations used for different types of accelerators, see footnote in Tab.1.

## References

- [1] R.W. Hamm, M.E. Hamm, Physics Today, June 2011, p.46
- [2] R. Hellborg, H.J. Whitlow, Y. Zhang (eds.), Ion Beams in Nanoscience and Technology, Springer, 2009
- [3] R.W. Hamm, M.E. Hamm (eds.), Industrial Accelerators and their Applications, World Scientific (2012)
- [4] R. Hellborg (ed.), Electrostatic Accelerators – Fundamentals and Applications, Springer, 2005
- [5] IAEA-TECHDOC-1560, Intern. Atomic Energy Agency, Vienna (2007); K. Noda, CERN Courier 50:5, 22-24 (2010)
- [6] Introduction to ion beam biotechnology, Trans. L. Yu, T. Vilaithong, I. Brown, Springer, 2006
- [7] IAEA Technical report Series 468 (2009)
- [8] R. Hellborg, G. Skog, Mass Spec Rev. 27:5, 398 (2008)

- [9] S.A.E. Johansson, J.L. Campbell, K.G. Malmqvist, Particle-Induced X-ray Emission Spectrometry (PIXE), Wiley, 1995
- [10] D.E. Power et al, Welding Journal, 79-2, 35 (2000)
- [11] B.R.R. Persson, Radiation Therapy, in Radiation – at Home, Outdoors and in the Workplace (eds. D. Brune, R. Hellborg, B.R.R. Persson and R. Pääkkönen), Scandinavian Sci. Pub., Oslo 2001
- [12] K. Nasta, C.-C. Kao, Synchrotron Radiation News, 20-4, 7 (2007)
- [13] R.H. Bossi et al (eds.), Non-destructive Testing Handbook, Vol.4, American Soc. for Non-destr. Testing, Columbus, OH (2000)

### 1.6.10 Laser, Wakefield and Plasma Accelerators

*J. Rosenzweig, UCLA*

The rf linac has evolved continuously from its birth after the World War II, an innovation made possible by the development of high power microwave sources for radar. The success of these devices is undeniable; however, with the insistent demand for increasingly high energy, both the size and cost of these “big science” machines has reached the billion-dollar level. The problem of miniaturizing the linac has been recognized and has stimulated research aimed at reinventing the accelerator scheme based on new physical principles that use lasers, plasmas and wakefields (advanced acceleration techniques) [1].

Present rf linacs are limited in acceleration gradient to  $\lesssim 100$  MV/m by two effects: first, the limit on maximum power and stored energy available and, second, by structure breakdown. To reach present limits, power and field energy considerations have yielded an optimization at shorter rf wavelength  $\lambda_{EM}$ , in X-band. Advanced accelerators aim to enhance field gradients to  $> \text{GV/m}$  (allowing a TeV machine in a km), arguing for diminishing  $\lambda_{EM}$  by at least an order of magnitude. In this frequency regime, metals are too lossy to be of use; the present approach to accelerator design must be abandoned in favor of new, more general “structures”. Further, to obtain high powers at short  $\lambda_{EM}$  new sources must be considered.

The most obvious coherent EM power source available is the laser, which easily produces copious infrared-to-optical power  $P$ , to  $> \text{TW}$ . With such high  $P$ , the paradigm for accelerating structures may be changed. Axisymmetry, essential

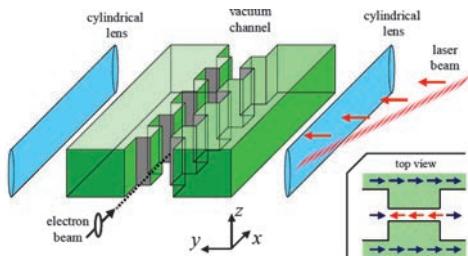


Figure 1: Schematic of laser-structure accelerator, showing laser injection and asymmetric (flat) accelerating beam, from [2].

when  $P$  is a limiting consideration, can be relaxed in favor of structures that closely resemble laser resonators [2], as in Fig.1. These structures, which rotate the laser’s transverse electric field to give longitudinal acceleration, are fabricated of low loss dielectric that gives breakdown limits in the GV/m range. To exploit higher fields, the particle trajectory may also be bent to have a periodic transverse component, as in the inverse free-electron laser (IFEL).

Dielectric laser accelerator (DLA) structures have been developed in a wide variety of designs, many inspired by optical sciences, e.g. photonic band-gap structures. Such devices have undergone testing at SLAC, with DLA-based acceleration demonstrated [3]. These studies have recently concentrated on pre-bunched beam injection, with the optical-scale bunching provided by an upstream IFEL. The IFEL interaction for micro-bunching is now a common tool for injection into optical accelerators. IFEL staging was shown at BNL [4]; two stages of IFEL were synchronized, the first giving micro-bunching and the next accelerating the beam with small energy spread. High energy gain was shown later at UCLA [5], with 14.5 MeV injected beam accelerated to  $\sim 35$  MeV peak in 25 cm. This experiment also displayed the higher harmonic IFEL interaction, which has extended the utility of IFEL for micro-bunching [6].

While the IFEL has shown more experimental progress than the DLA, its application to high energy physics accelerators is limited by synchrotron radiation. It is now under study, however, for sub-GeV applications, such as inverse-Compton scattering (ICS) production of MeV photons for nuclear materials identification. On the other hand, DLAs promise more control and

flexibility than IFELs, and scale to high-energy application well. Indeed, with a structure, an optical undulator [7] may be built that permits an x-ray FEL using  $< 500$  MeV beams. These may be most attractively obtained from a laser structure or IFEL accelerator; this combination is now under study to give compact, all optical, coherent x-ray source, i.e. the GALAXIE project, a UCLA-SLAC-BNL-Penn State consortium.

DLA research presents many challenges. As many schemes utilize geometries that are sub- $\lambda_{EM}$  in one transverse direction, the normalized emittance in this direction must be very small, in the  $\epsilon_N \sim 10^{-9}$  m-rad range. Further, the longitudinal dynamics is very challenging, as relevant time scales are 4 orders of magnitude smaller than in current devices. Thus, while much shorter  $\lambda_{EM}$  is still demanded, options midway from rf to optical, in the mm-THz regime, are attractive. Here, the dynamics requirements are relaxed, and a key breakdown mechanism in dielectrics, avalanche ionization, is mitigated. However, there is no readily available EM power source that can generate  $E_z > \text{GV/m}$  in a THz-scale structure. This need has stimulated the invention of *wakefield* acceleration, in which the EM fields of the accelerating wave are created inside of the structure itself by an intense, relativistic particle beam. This drive beam may be of lower quality and energy than a trailing, accelerating beam. Further, the drive beam may be specially shaped (in, e.g. a rising triangular current profile) to give much larger acceleration in the trailing beam than deceleration in the driver; this relationship is termed the *transformer ratio*. When one uses a dielectric structure in such a scheme (Fig.2), it is called a dielectric wakefield accelerator (DWA).

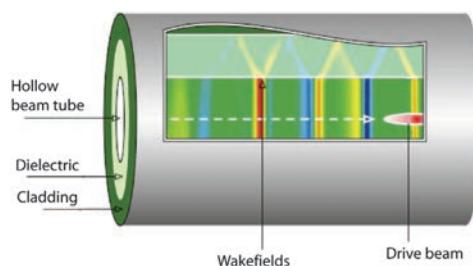


Figure 2: Schematic of dielectric wakefield accelerator showing drive beam hollow dielectric tube generating Cerenkov wakes.

The first DWA experiments in the high frequency rf regime ( $\sim 30$  GHz) at ANL [8] achieved field gradients up to 100 MV/m. The field scaling of the coherent Cerenkov radiation (CCR) emitted in a DWA is  $N_b\omega^2$ , where  $N_b$  is the number of beam charges and  $\omega$  is the EM frequency. To excite a high frequency mode, the drive beam pulse (or train) must have significant Fourier components at this frequency. With sub-100 fs beams available in recent years, GV/m DWA fields became accessible. First measurements at SLAC FFTB have shown that  $> 5$  GV/m fields can be excited in a SiO<sub>2</sub> DWA before breakdown [9]. Subsequent experiments at UCLA have demonstrated that the CCR exiting a DWA tube is quite narrow band [10], yielding a unique high power THz source for frequency-specific applications.

DWA research is continuing in many laboratories: the ANL AWA, BNL ATF, and the nascent SLAC FACET facility. These programs will explore breakdown mechanisms, resonant multi-pulse excitation, acceleration in multi-GV wakes, enhanced transformer ratios, and are now developing asymmetric (slab) structures [11]. The use of slab structure with asymmetric beams permits acceleration of higher charge trailing beams, and also strongly suppresses transverse wakefields [12]; these advantages are shared with many optical accelerator structures.

Ultimately, both the DLA and DWA are limited by breakdown. For  $E_z > 5$  GV/m, one must consider the accelerating “structure” to be broken down, that is to be plasma. The scenario in which a drive beam excites plasma waves that support large  $E_z$  is termed the plasma wakefield accelerator (PWFA) [13]. In the linear regime, where the plasma density  $n_0$  greatly exceeds the beam density  $n_b$ , the wake wave is an electrostatic oscillation with plasma frequency  $\omega_p = k_pc = \sqrt{4\pi e^2 n_0/m_e}$ . As long as the wake is efficiently excited ( $k_p\sigma_z < 1$ ,  $k_p\sigma_r < 1$ ), the excited field  $E_z \cong N_b e^2 k_p^2 / 2 \sim (n_b/n_0) E_{WB}$ , where the wave-breaking field  $E_{WB} \approx 0.96 \sqrt{n_0(\text{cm}^{-3})}$  V/cm. When  $n_b \approx n_0$   $E_z$  approaches  $E_{WB}$ ; for a modest plasma density of  $10^{18} \text{ cm}^{-3}$ , this is  $\sim 100$  GV/m.

The first PWFA was performed at ANL [14]. Simultaneously, it was proposed to use plasma wakefields for creating lenses (Sec.7.3.13) with extreme focusing strength for a linear collider final focus. The case with  $k_p\sigma_z > 1$ ,  $k_p\sigma_r < 1$  and  $n_b > n_0$  (underdense plasma lens) is most attractive, producing nearly aberration-free focusing in

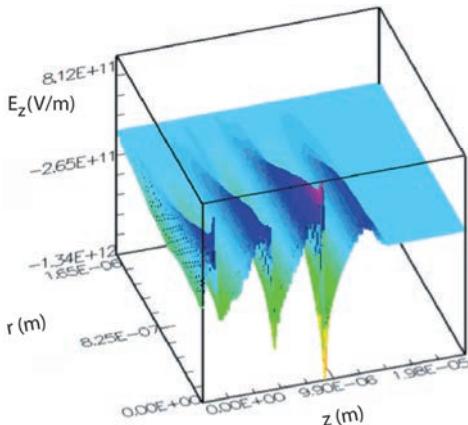


Figure 3: PWFA field  $E_z$  excited by 2 fs,  $\sigma_x = 77$  nm, 20 pC beam, with  $n_0 \approx 8 \times 10^{19}$  cm $^{-3}$ .

an electron-evacuated ion column, with fields linear in  $r$  and independent of  $\zeta = z - v_b t$ . For the example of a  $n_0 = 10^{18}$  cm $^{-3}$  plasma the equivalent quad strength in the lens is  $2\pi e n_0 = 30$  MT/m,  $> 5$  orders of magnitude stronger than a conventional magnet. Overdense plasma ( $n_b < n_0$ ) lensing was demonstrated at UCLA [15]; underdense plasma lensing has been observed at FNAL [16].

Indeed, PWFA operation with  $n_b \gg n_0$  (and  $k_p \sigma_z < 1$ ,  $k_p \sigma_r < 1$ ) where electrons are expelled from the beam channel has considerable advantages [17]. In addition to the linear ion-column focusing inside the blowout “bubble” region, as the currents supporting the (pure EM) wave lay outside of the bubble,  $E_z$  (see Fig.3) is independent of  $r$  just as in a standard rf linac. Ion column focusing in the blowout regime was first shown at ANL [18], and acceleration at  $> 100$  MeV/m observed at FNAL. With the advent of short ( $\sigma_z \approx 20$   $\mu$ m), high charge (3 nC) beams at the SLAC FFTB, peak  $E_z$  in the blowout regime at 10–100 GV/m could be reached [19]. These studies demonstrated a wide variety of effects, including betatron-oscillation x-ray production, and a doubling of the beam energy at the drive beam tail [20].

New experimental work in PWFA at FACET has begun, along with a DWA program. It will explore wakefield acceleration with positrons, as well the creation of ramped beams and separate

witness beams. With even shorter beams available at the LCLS ( $\sigma_z \approx 0.6$   $\mu$ m,  $Q = 20$  pC) a blowout regime PWFA with  $E_z > 1$  TV/m may be created (Fig.3). This option is under study [21].

While the PWFA has yet to produce a low energy spread and low  $\epsilon_N$  beam, the laser-driven version of the scheme (laser wakefield accelerator, LWFA) has done so, trapping and accelerating plasma electrons to yield  $\epsilon_N \sim$ mm-mrad,  $>$ kA beams up to  $\sim$ 1 GeV [22]. With some improvements in beam quality, these beams may create a “table-top” FEL [23]. Further information on LWFA and other laser-driven plasma accelerators is found in Sec.7.3.13 and [24].

## References

- [1] See, e.g. the Advanced Accelerator Concepts series proceedings, the last being *14th Advanced Accelerator Concepts Workshop*, Eds. S.H. Gold, G.S. Nusinovich, AIP Proc. 1299 (2010)
- [2] T. Plettner et al, PRST-AB 9, 111301 (2006); J.B. Rosenzweig, A. Murokh, C. Pellegrini, PRL 74, 2467 (1995)
- [3] T. Plettner et al, PRL 95, 134801 (2005)
- [4] W.D. Kimura et al, PRL 92, 054801 (2004)
- [5] P. Musumeci et al, PRL 94, 154801 (2005)
- [6] C.M.S. Sears et al, PRL 95, 194801 (2005)
- [7] T. Pletter, R. Byer, PRST-AB 11, 030704 (2008)
- [8] W. Gai et al, PRL 61, 2756 (1988)
- [9] M.C. Thompson, PRL 100, 21 (2008)
- [10] A.M. Cook et al, PRL 103, 095003 (2009)
- [11] G. Andonian et al, PRL 108, 244801 (2012)
- [12] A. Tremaine, J. Rosenzweig, P. Schoessow, PR E56, 7204 (1997)
- [13] P. Chen, PA 20, 171 (1985)
- [14] J.B. Rosenzweig et al, PRL 61, 98 (1988)
- [15] G. Hairapetian et al, PRL 72, 1244 (1994)
- [16] M.C. Thompson et al, Phys. Plasmas 17, 073105 (2010)
- [17] J.B. Rosenzweig et al, PRA – Rapid Comm. 44, R6189 (1991)
- [18] N. Barov et al, PRL 80, 81 (1998)
- [19] M.J. Hogan et al, PRL 95, 054802 (2005)
- [20] I. Blumenfeld et al, Nature 741, 445 (2007)
- [21] J.B. Rosenzweig et al, NIM A, doi:10.1016/j.nima.2011.01.073
- [22] W.P. Leemans et al, Nature Physics 2, 696 (2006)
- [23] M. Fuchs et al, Nature Physics 5, 826 (2009)
- [24] E. Esarey, C.B. Schroeder, W.P. Leemans, Rev. Mod. Phys. 81, 1229 (2009)

### 1.6.11 Linear Accelerators for Electrons

G.A. Loew, SLAC

Electron (positron) linear accelerators (linacs) are axially rectilinear structures which capture a beam from an electron injector (positron source), and accelerate it to a desired energy [1, 2, 3]. These linacs rely on rf (typically 0.5-30 GHz) energy to produce the accelerating electric field. Above a few MeV, dc machines no longer work because cumulative high voltages are difficult to obtain.

The rf energy is used to launch a traveling wave (TW) or a standing wave (SW) in an array of microwave cavities or cells (Fig.1). (Sects.1.6.12, 2.4.6, 7.3.6). In both cases, the structure is designed so that the phase of the wave is synchronized with the beam. Because of the small rest mass of  $e^\pm$ , above a few MeV, the needed rf phase velocity  $v_p \approx c$ . Generally, when entering the linac, particles are already considerably bunched, and what little compression remains to be done occurs during the first few MeV of acceleration (Fig.2). The bunches can then ride at a constant rf phase and corresponding acceleration field.

The structures shown in Fig.1 are periodic (or quasi-periodic, see below) in that their cell geometry is invariant with respect to longitudinal translation. The periodic loading is necessary because a smooth waveguide would have  $v_p > c$ . The irises or other periodic obstacles create an infinite family of space-harmonic modes (Fig.3). The fundamental mode ( $n = 0$ ) generally has the largest amplitude and is used for acceleration. The Brillouin diagram [4], Fig.3, corresponds to a TW structure in which this fundamental mode has  $v_p = c$  at a phase shift of  $2\pi/3$  per cell. Fig.3

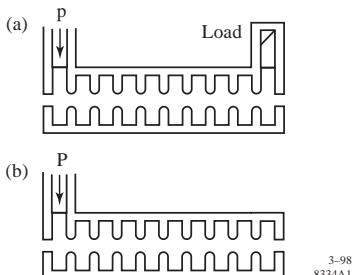


Figure 1: (a) Traveling-wave structure with matching input iris and matched load at output. (b) Standing-wave structure with total reflection at output end and matching input iris (steady state).

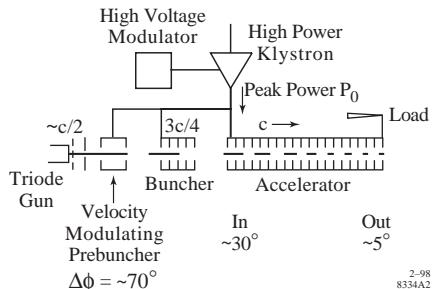


Figure 2: Schematic of a single-section, single-power-source linac, with typical injector.  $\Delta\phi$  is bunch length in units of rf phase. The gridded triode gun is sometimes replaced by a laser-driven photocathode or an rf gun where the cathode is embedded inside an rf cavity. Progressive bunching is graphically illustrated. Multi-section linacs are simply constructed by adding sections linearly beyond the first one.

exhibits a second upper-branch (there are an infinity of such higher-order modes, HOM) which intercepts the  $v_p = c$  line at a point with negative slope (backward-wave HEM<sub>11</sub> mode). These modes are commonly called wakefields when they are excited by a beam bunch. The wakefields then cause particles later in this bunch or subsequent bunches to be deflected transversely, producing emittance growth and eventually beam break-up and particle loss. This problem can be controlled by appropriate solenoidal focusing (at low energy) and quadrupole focusing (at higher energy), or by cell detuning (see below).

Most linac structures are made out of high quality copper which, however, has a finite loss which causes attenuation. For a given amount of pulsed rf peak power  $P_0$  injected into an accelerator section, it is not economical to extend its length  $\ell$  beyond the point where  $\sim 70\%$  of the power has been dissipated. In a TW structure, the remaining power is dumped into a load. The input is matched so that there is no reflection at the source. In a SW structure, there are also losses and while some power is reflected from the end, it is possible to match the input so that there is no power reflection in steady state. Because a standing wave can be decomposed into two oppositely moving traveling waves, the only efficient way of operating a SW structure is in a mode with a  $\pi$ -phase shift per cell on axis, which means that both traveling waves are confluent and produce acceleration (in reality, for mode stability reasons, the

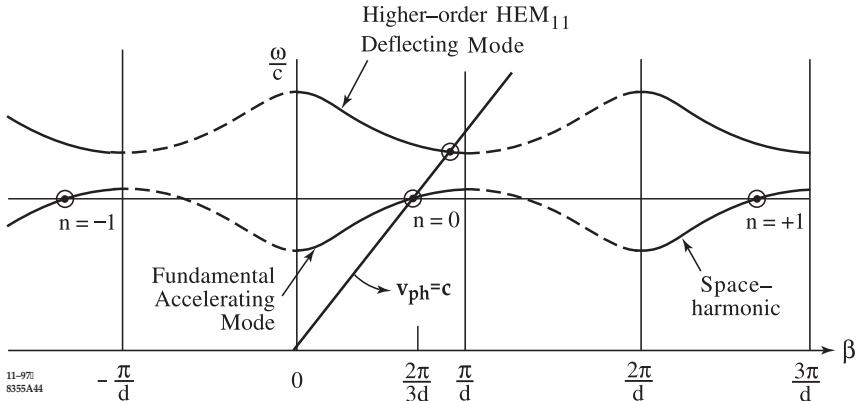


Figure 3: Typical Brillouin diagram for a disk-loaded waveguide, showing fundamental accelerating mode operating at  $2\pi/3$  phase shift per cell, and one branch of a higher-order HEM<sub>11</sub> deflecting mode.

actual phase-shift per cell in e.g. the SW side-coupled structure invented at Los Alamos [5] is  $\pi/2$ ). Advantages and disadvantages of TW vs. SW structures are discussed in [5].

All periodic linac structures are characterized by a figure of merit called the shunt impedance per unit length,  $r$ . The steady-state no-load energy acquired by a particle riding on top of the synchronous fundamental wave is

$$V = K \sqrt{P_0 \ell r} \quad (1)$$

where the proportionality constant  $K < 1$  for a TW structure depends on the attenuation of the section, and  $K = 1$  for the perfectly matched SW structure, assuming no appreciable field reduction due to beam loading. If the linac consists of  $N$  identical sections, then the total energy gain is  $NV$ . With a peak beam current  $I$ , the steady-state energy (i.e., after the appropriate filling times) is reduced by a subtractive term  $\propto rIl$ .

TW structures may be designed to be quasi-periodic. Two examples stand out. The first is the *constant-gradient* structure in which dimensions are tapered so as to decrease the group velocity linearly with length. The fields are thereby caused to be constant as opposed to exponentially decaying as a function of length, as is the case in regular *constant-impedance* structures. The second one is a variation of the former where the cavity dimensions are varied so as to keep the accelerating fields approximately constant while giving the HOM frequencies a Gaussian distribution which causes them to decohere rapidly in time after their excitation by the beam. A further variation of this

scheme is to equip every cell with four symmetrical side-openings which enable the HOMs with both horizontal and vertical polarizations to leak out into manifolds while leaving the fundamental accelerating mode undisturbed [6].

While the accelerator structure is the heart of the linac, there are other essential components and subsystems. Fig.2 shows a high power rf source which in most machines is a klystron (magnetrons are used in single-section low energy machines) with its associated high voltage modulator. Because of the high peak power required (typically 4-80 MW), these klystrons have low duty cycles ( $\sim 100$  Hz) and pulse lengths  $\sim$  a few  $\mu s$ . To increase the peak power, modern linacs sometimes use rf pulse compression systems (Sec.6.7), which temporarily accumulate the rf energy in a storage device, thereby enhancing the peak power emitted in a pulse at the expense of its width. Fig.2 also shows an injector with its triode gun, pre-buncher and buncher. Other important subsystems include rf drive and phasing, focusing, vacuum, alignment, water cooling, ac power, instrumentation and control [3].

For a given total energy, the length of a linac can be reduced by increasing its accelerating gradient. Typical machines use gradients from a few MV/m up to proposed  $\sim 100$  MV/m. At low gradients, the structure can sometimes suffer from a parasitic resonant phenomenon (multipactor) in which a surface-emitted electron gets accelerated, hits the surface elsewhere (or returns to the same point), and then ejects new secondary electrons which produce an avalanche condition (Sec.6.9).

## Sec.1.6: GLOSSARY OF ACCELERATOR TYPES

Table 1: Linac parameters for SLC (achieved) [7], NLC (design) [8], and LCLS (design) [9].

|                           | SLC   | NLC            | LCLS                |
|---------------------------|-------|----------------|---------------------|
| Energy (GeV)              | 50    | $2 \times 250$ | 15                  |
| Gradient(MeV/m)           | 17    | 57             | 17                  |
| $f_{rf}$ (GHz)            | 2.856 | 11.424         | 2.856               |
| klystron                  |       |                |                     |
| #/linac                   | 224   | 3270           | 70                  |
| peak power(MW)            | 65    | 75             | 65                  |
| pulse length( $\mu$ s)    | 3.5   | 1.5            | 3.5                 |
| $e^-/bunch (10^{10})$     | 4     | 1.0            | 0.63                |
| $\epsilon_{Nx}$ (mm-mrad) | 45    | 4              | 1.0                 |
| $\epsilon_{Ny}$ (mm-mrad) | 5     | 0.04           | 1.0                 |
| $\sigma_z$ (mm)           | 1.0   | 0.15           | 0.03 <sup>[a]</sup> |
| $\sigma_\delta (10^{-3})$ | 1.5   | 3.0            | < 1 <sup>[a]</sup>  |

<sup>[a]</sup> after full compression.

At high fields, electron field emission takes place, which can have a variety of deleterious effects: parasitic absorption of energy, wakefields, dark-current producing spurious radiation and backgrounds, and eventually rf breakdown which can make the linac inoperative. Field emission is caused by a variety of surface irregularities, impurities, dust, and contamination, and can be controlled, up to a point, by proper surface fabrication, treatment and cleaning (Sec.6.13). Another type of structure damage more recently discovered at high magnetic fields is excessive rf pulse heating resulting in copper melting.

Tab.1 gives the parameters achievable by the full SLAC  $e^-$  linac, as well as originally proposed design parameters for the NLC and the LCLS.

Electron and positron linacs are used for many purposes [10]. The largest number of electron linacs (many thousands) are radiotherapy machines which can be found in hospitals worldwide (energy  $\sim$ 4-30 MeV, x-rays produced by electrons impinging on targets, or direct electron radiation, Sec.1.6.14). Low energy linacs are also used in industry for sterilization of various materials and products, x-ray radiography, etc. (Sec.1.6.9). Most of these linacs consist of single sections and power sources. Linacs of higher energy (50 MeV-50 GeV) are used in laboratories for nuclear and particle physics and as injectors into  $e^\pm$  storage rings of various types [11]. As this article is being written, the international particle physics community is still assessing the possibility of building an  $e^+e^-$  linear collider in the TeV CM energy range. Two main designs, one (ILC) using

an rf frequency of 1.3 GHz with superconducting technology, and the other (CLIC) using 12 GHz with room temperature technology, are still being considered, among the many options proposed in 2003 [12].

## References

- [1] Linear Accelerators, ed., A. Septier, P.M. Laposoltte, North-Holland (1970)
- [2] G.A. Loew, R. Talman, AIP Conf. Proc. 105 (1983) p.1
- [3] R.B. Neal et al, The Stanford Two-Mile Linear Accelerator, Benjamin (1968)
- [4] L. Brillouin, Wave Propagation in Periodic Structures, Dover (1953)
- [5] R.H. Miller, Proc. Linear Acc. Conf., SLAC-303 (1986) p.200
- [6] R.H. Miller et al, Proc. Linear Acc. Conf. (1996)
- [7] P. Emma, PAC 95, p.606
- [8] Zeroth Order Design for Next Linear Collider, SLAC-474 (1996)
- [9] V.K. Bharadwaj, PAC 97
- [10] G.A. Loew, Proc. Proton Linear Acc. Conf. (1976) p.217
- [11] J. Clendenin et al, Compendium of Scientific Linacs (1996)
- [12] G.A. Loew et al, Int. Linear Collider Tech. Rev. Comm. Second Report, SLAC-R-606 (2003)

### 1.6.12 Linear Accelerators for Protons

S. Henderson, FNAL  
A. Aleksandrov, ORNL

First demonstrated in 1947 by Alvarez and co-workers at Berkeley [1], proton linear accelerators [2, 3] find use as dedicated linacs for nuclear physics, injectors for high-energy synchrotrons for nuclear and particle physics, and drivers for spallation neutron production. Often,  $H^-$  ions (Sec.7.1.4) are accelerated instead of protons to make use of charge-exchange injection in a downstream synchrotron or accumulator ring. Tab.1 summarizes parameters of existing proton/ $H^-$  linear accelerators. Beam energies extend to 1 GeV and duty factors to several percent. Peak beam currents reach  $\sim$ 200 mA for protons and  $\sim$ 50 mA for  $H^-$  beams, with pulse lengths ranging from  $\sim$ 10  $\mu$ s to 1 ms.

Proton linacs generally utilize RFQs (Sec.1.6.18) or Cockcroft-Waltons (Sec.1.6.7) as injectors, and therefore provide acceleration for particle velocities of  $\beta \simeq 0.03$  to  $\beta \simeq 0.9$ . Acceleration is provided by TM<sub>010</sub>-like fields (Sec.6.6)

Table 1: Parameters of operating proton/H<sup>-</sup> linacs (taken from [4] except where noted).

| Facility             | Species          | Design Energy [MeV] | Structure (energy [MeV])     | RF Frequency [MHz]       | Peak Beam Current [mA]       | Beam Pulse Length [ $\mu$ s] | Rep Rate [Hz] |
|----------------------|------------------|---------------------|------------------------------|--------------------------|------------------------------|------------------------------|---------------|
| IPNS/ANL [5]         | H <sup>-</sup>   | 50                  | DTL                          | 200.06                   | 12                           | 90                           | 30            |
| FNAL [6]             | H <sup>-</sup>   | 400                 | DTL (116)                    | 201.25                   | 50                           | 60                           | 15            |
|                      |                  |                     | CCL                          | 805                      |                              |                              |               |
| Indiana U.           | H <sup>-</sup>   | 7                   | DTL                          | 425                      | 1                            | 400                          | 5             |
| LANSCE/LANL          | p/H <sup>-</sup> | 800                 | DTL (100)                    | 201.25                   | 17(p)<br>11(H <sup>-</sup> ) | 825                          | 120           |
|                      |                  |                     | CCL                          | 805.0                    |                              |                              |               |
| BNL [7]              | H <sup>-</sup>   | 200                 | DTL                          | 201.25                   | 37                           | 530                          | 7.5           |
| IHEP/Beijing         | p                | 35                  | DTL                          | 201.25                   | 40                           | 150                          | 12.5          |
| KEK                  | H <sup>-</sup>   | 40                  | DTL                          | 201.07                   | 18                           | 80                           | 20            |
| Kyoto U.             | p                | 7                   | DTL                          | 433                      | 0.6                          | 50                           | 180           |
| DESY                 | H <sup>-</sup>   | 50                  | DTL                          | 202.56                   | 20                           | 30                           | 1             |
| MMF/INR              | p/H <sup>-</sup> | 600                 | DTL (100)                    | 198.2                    | 50                           | 80                           | 100           |
|                      |                  |                     | DAW                          | 991.0                    |                              |                              |               |
| I-2/ITEP             | p                | 25                  | DTL                          | 297                      | 200                          | 20                           | 2             |
| ISTRRA/ITEP          | p                | 36                  | DTL                          | 297                      | 150                          | 10                           | 1             |
| I-100/Protvino       | p                | 103                 | DTL                          | 148.5                    | 100                          | 100                          | 1             |
| URAL-30/Protvino [8] | p                | 30                  | RFQ-DTL                      | 148.5                    | 100                          | 10                           | 16.6          |
| Linac 2/CERN         | p                | 50                  | DTL                          | 202.56                   | 180                          | 120                          | 2             |
| ISIS/RAL             | H <sup>-</sup>   | 70                  | DTL                          | 202.5                    | 25                           | 500                          | 50            |
| J-PARC [9]           | H <sup>-</sup>   | 400                 | DTL (50)<br>SDTL (191)       | 324<br>324<br>ACS<br>972 | 50                           | 500                          | 50            |
| SNS/ORNL[10]         | H <sup>-</sup>   | 1000                | DTL (87)<br>CCL (186)<br>SCL | 402.5<br>805.0<br>805.0  | 38                           | 1000                         | 60            |

established in accelerating gaps (or cells) which are arranged in standing-wave multicell cavities of two types. In the first, a long structure with tens or even hundreds of cells is excited by a single rf power source. The physical spacing between cells varies along the length of the structure in accordance with the design particle velocity profile. In the second, the distance between cavities is fixed, and relative beam/rf phase is adjusted by varying the rf generator phase.

The principal challenge in modern high-intensity proton linac design and operation is the minimization of beam loss due to halo growth [11, 12, 13]. An often stated loss criteria [14, 15] demands particle loss rates less than 1 watt of beam power per meter in order to limit residual activation dose rates to less than  $\sim 100$  mrem/hr at 30 cm after 4 hr cooldown. See Sec.2.4.6 for a description of the beam dynamics in proton linacs.

### Rf structures

Drift tube linac The Drift Tube Linac (DTL) (Sec.7.3.5), is the most common structure for acceleration of proton/H<sup>-</sup> beams in the velocity range  $0.03 < \beta < 0.4$  ( $\sim 0.5$  MeV to  $\sim 100$  MeV) and so constitutes the low-energy portion of nearly all proton linacs. DTLs generally operate in the frequency range  $\sim 200$ -400 MHz. A full DTL system consists of one or more cylindrical pillbox resonant tanks, each excited in a TM<sub>010</sub>-like standing wave mode. Drift tubes, placed with a nominal spacing of  $\beta\lambda$ , provide a field-free drift region during the decelerating portion of the rf cycle. The accelerating field in each gap oscillates at the same phase and frequency (the zero-mode). The length of cell  $i$  is

$$l_i = \left( \frac{\phi_{s,i+1} - \phi_{s,i}}{2\pi} + 1 \right) \beta_i \lambda \quad (1)$$

where  $\phi_{s,i}$  is the synchronous phase for cell  $i$  and  $\beta_i$  is the velocity at the exit of the cell.

The accelerating gradient is usually constant from cell to cell, although a field ramp is sometimes included to maintain smooth longitudinal focusing. The synchronous phase at each cell (the “phase law”) is selected to provide adequate longitudinal focusing (typically between  $-30^\circ$  and  $-20^\circ$ ) and may also be ramped to adjust phase acceptance, or for longitudinal matching. Either permanent magnet or electromagnetic quadrupoles are contained within drift tubes to provide transverse alternating gradient focusing.

At higher energy, the effective shunt impedance begins to decrease, so that the DTL structure begins to become less efficient than other structures above  $\sim 100\text{-}200$  MeV.

Coupled-cavity structures The various “high-energy” rf structures exploit the remarkable properties of the  $\pi/2$  mode of excitation in a biperiodic array of coupled resonant cavities [16]. In the  $\pi/2$  mode, every other cavity is excited, while adjacent cavities are unexcited, and therefore do not contribute to beam acceleration. The  $\pi/2$  mode has the following properties: i) the field amplitude in excited cells is independent of cell frequency errors in first order, therefore making the structure less sensitive to construction tolerances and tuning errors; ii) rf losses do not produce a phase shift in excited cavities; and iii) rf losses produce a cavity amplitude decrease only in second-order.

Efficient acceleration is achieved by placing the nominally unexcited cells off the beam axis, so that the beam traverses every other cell in the long array of coupled cavities. The relative rf phase shift between accelerating cells is therefore  $\pi$ , and the (excited) cell spacing is nominally  $\beta\lambda/2$ . A single klystron powers a long coupled cavity linac (CCL) module, which consists of many multi-cell tanks or segments (each of which may have  $\sim 10$  accelerating cells) forming a long chain of as many as 200 coupled oscillators. Segments are coupled one to the next by off-axis “bridge couplers” that span the intercavity drift spaces where quadrupoles and diagnostic devices are located.

A high shunt impedance at higher velocity ( $\beta > 0.4$ ) makes these structures useful for acceleration of a beam delivered by a DTL. The most common coupled cavity structure is the side-coupled structure used at the LANSCE linac [16], the FNAL linac [17] and the SNS linac [18], all of which operate at 805 MHz, and which follow a  $\sim 100$  MeV DTL. The Disk and Washer structure

is in use at the Moscow Meson Factory [19]. An Annular Coupled Structure (ACS) is being built for J-PARC [20].

Elliptical superconducting cavities The SNS utilizes elliptical SC multicell cavities (Sec.7.3.10) with geometric-beta (the synchronous particle velocity)  $\beta_g < 1$ , for acceleration from 186 to 1000 MeV [21, 22]. Two structures with  $\beta_g = 0.61$  and  $\beta_g = 0.81$  provide acceleration in the velocity range  $0.55 < \beta < 0.90$ . The cavity is excited in the  $\pi$  mode with cell spacing equal to  $\beta_g\lambda/2$ . The benefits of SC cavities realized in other applications, namely, high accelerating gradients, compact designs, large aperture, lower rf power requirements, etc., are gained in application to proton/ $H^-$  linacs as well.

Other structures A large number of other structure types have been studied and several have been built. These include the CCDTL structure [23], and the crossed-bar H-cavities (CH-cavity) under study for the FAIR project [24]. A separated drift tube linac structure (SDTL) consisting of short five-cell tanks with quadrupoles between tanks, is used at the J-PARC facility for acceleration from 50-190 MeV [25]. A DTL with space-periodic RFQ focusing (a RFQ-DTL) operates at IHEP Protvino [8]. Additionally, SC spoke resonators are under consideration for the low-energy portion of future high-power proton linacs [26].

Rf power considerations For a shunt impedance defined by  $R_s = V_0^2/P$ , the shunt impedance per unit length is  $Z = E_0^2/(P/L)$  and the effective shunt impedance per unit length is  $ZT^2 = (E_0T)^2/(P/L)$ , where  $E_0$  is the peak axial electric field,  $T$  is the transit-time factor and  $P$  is the rf power. For example, SNS DTL tank 6 (length 6.34 m), has  $ZT^2 = 39$  M $\Omega$ /m,  $E_0T = 2.8$  MV/m, and so requires an *rf structure power* of 1.25 MW to establish the field.

Design codes The most commonly used design codes for rf cavities and proton linac structures are the POISSON/SUPERFISH codes [27] for electromagnetic field calculation, and PARMILA [28] for generating the linac layout and performing beam dynamics computations.

**Rf setpoint determination** To ensure good quality of the accelerated beam, the phase and amplitude of the rf cavities must be set and maintained very close to their design values (Sec.2.4.6). Typical requirements are less than

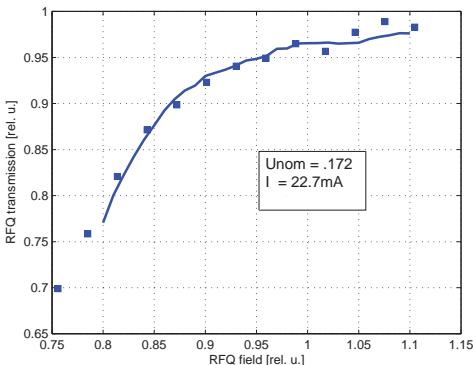


Figure 1: Measured transmission vs rf field amplitude for the SNS RFQ compared to a model prediction.

$1^\circ$  phase error, and less than 1% amplitude error. Beam-based tuning techniques that make use of model-based methods are best suited.

RFQ tuning RFQ tuning requires setting the rf field amplitude only. A fit of the measured beam transmission vs rf power to a curve calculated from the model allows the correct set point to be determined. An example is shown in Fig.1 [29].

DTL tuning The rf phase and amplitude must be set for each DTL tank. One method, *the acceptance scan*, determines set points by comparing the measured beam transmission vs rf phase to the *phase width* predicted by a model (Sec.2.4.6). Low-energy particles are absorbed in an energy degrader, so that only properly accelerated beam current is measured with a faraday cup. An example SNS DTL acceptance scan is shown in Fig.2. A second method, *phase scan signature matching* [30], makes use of the measured beam phase at one or two locations downstream of the tank vs the tank rf phase and amplitude. The rf set points and energy of the incoming beam are obtained by comparing the measured curve to the model. An example is shown in Fig.3 [31].

CCL tuning The rf phase and amplitude have to be set for each CCL module. In the *Delta-T scan* method, the beam phase at two locations downstream of the module is measured vs the module rf phase with the rf on and off. The set points are calculated using the measured data and coefficients derived from the model [32, 33]. Phase scan signature matching analogous to that described for the DTL can also be used.

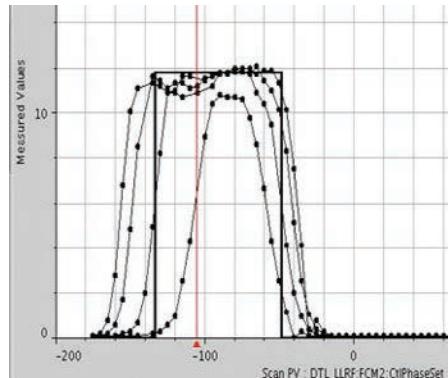


Figure 2: Measured beam current vs SNS DTL rf phase for several different DTL rf amplitudes. Curves with larger phase width correspond to higher rf field amplitude.

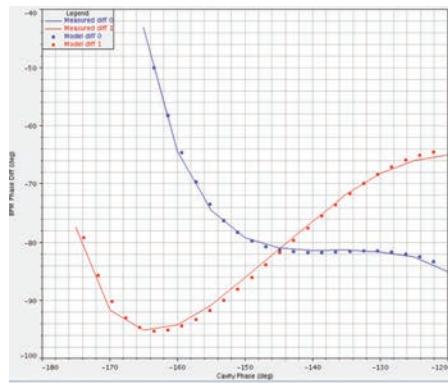


Figure 3: Measured difference in beam phase recorded by two beam phase monitors vs SNS DTL rf phase for two different DTL rf amplitudes. The solid curves show measured data and the points show results of the model-based fit.

SCL tuning The rf phase and amplitude have to be set for each cavity in the SCL. A phase scan analogous to the one described for the DTL is used to determine the input beam energy and phase, and the cavity field amplitude [34].

## References

- [1] L.W. Alvarez et al, RSI 26 (1955) 111
- [2] T. Wangler, Principles of RF Linear Accelerators, Wiley (1998)
- [3] P. Lapostolle, A. Septier, eds., Linear Accelerators, North-Holland (1970)

## Sec.1.6: GLOSSARY OF ACCELERATOR TYPES

- [4] J. Clendenin et al, CERN/PS 96-32(DI); <http://linac96.web.cern.ch/Linac96/Compendium/COMPENDI.PDF>
- [5] J.C. Dooling et al, PAC 01, p.1653
- [6] L. Allen et al, LINAC 96, p.329
- [7] J. Alessi et al, LINAC 96, p.773
- [8] Yu. A. Budanov et al, LINAC 04, p.285
- [9] Acc. Technical Design Report for J-PARC, JAERI-Tech 2003-44
- [10] N. Holtkamp, LINAC 04, p.837
- [11] Proc. 29th ICFA Adv. Beam Dyn. Wkshp, HALO 2003, AIP Conf Proc. 693
- [12] S. Nath et al, PAC 03, 1515
- [13] F. Gerigk et al, LINAC 02, p.569
- [14] 7th ICFA Mini-Workshop on High Intensity, High-Brightness Beams (1999)
- [15] N. Catalán-Lasheras et al, EPAC 02, p.1013
- [16] E.A. Knapp, in [3], p.601
- [17] C.W. Schmidt, PAC 93, p.1655
- [18] J.H. Billen, PAC 01, p.1104
- [19] V.G. Andreev et al, LINAC 1972, LA-5115, p.114
- [20] V.V. Paromonov, KEK Report 2001-14 (2001)
- [21] C. Rode, PAC 01, p.619
- [22] G. Ciavoti et al, PAC 01, p.484
- [23] J.H. Billen et al, LINAC 94, p.341
- [24] L. Groening et al, LINAC 04, p.42
- [25] Y. Yamazaki, LINAC 04, p.554
- [26] S.H. Kim, LINAC 04, p.549
- [27] POISSON/SUPERFISH, LANL LA-UR-96-1834
- [28] H. Takeda, LANL LA-UR-98-4478
- [29] A. Aleksandrov, PAC 05, p.97
- [30] T.L. Owens et al, PA 48 (1994) 169
- [31] J. Galambos et al, PAC 05, p.1491; J. Galambos, private communication
- [32] K. Crandall, LANL Report LA-6374-MS (1976)
- [33] A. Feschenko et al, PAC 05, p.3064
- [34] S. Henderson et al, PAC 05, p.3423

### 1.6.13 Livingston Chart *D. Hartill, Cornell U.*

The historical rise of the beam energy of accelerators when new technologies were developed to accelerate particle beams can be graphically illustrated by the Livingston Chart. The chart is named after Stanley Livingston who first used this method of showing this history. (Fig.1) [1]. The important parameter for new particle production and interactions is the center of mass energy which depends on the fixed target particle mass in addition to the beam energy. The chart in Fig.1 assigns an equivalent beam energy for the colliders

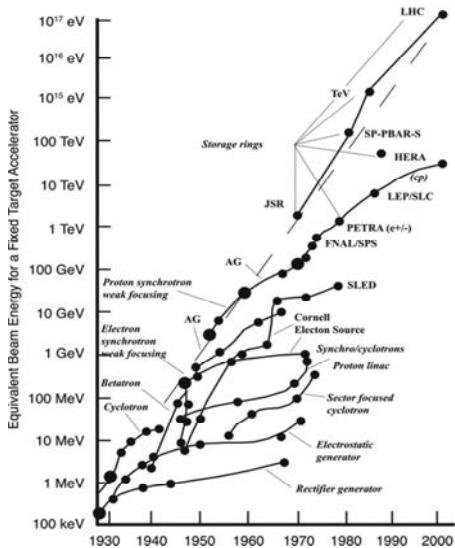


Figure 1: Livingston Chart.

where the center of mass energy is twice the beam energy for equal energy beams (Sec.1.6.3)

## References

- [1] M.S. Livingston, J.P. Blewett, Particle Accelerators, McGraw-Hill (1962)

### 1.6.14 Medical Applications of Accelerators

*J. Alonso, LBNL*

#### 1.6.14.1 Radiation therapy

Ionizing radiation in sufficient doses kills cells. As a result of much biological and clinical research, radiation is now highly effective for cancer therapy [1]-[4]. Precise tumor definition with CT and MRI imaging is crucial, as is dose-control to the percent level. Fig.1 shows energy-deposition characteristics of radiation modalities used, beam-delivery systems must work within these constraints to achieve 3-D conformation of dose to the tumor and avoid unacceptable damage to nearby healthy tissue.

**X-ray therapy** [5, 6] is the most widely used radiation treatment. The x-ray depth-dose relation is basically exponential. Treating a tumor deep inside a patient involves significant doses up- and down-stream of the treatment field. These doses can be mitigated by multi-port treatments in

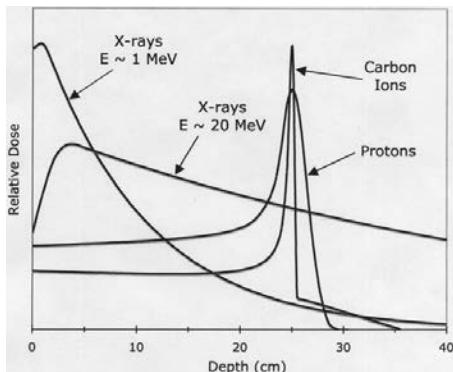


Figure 1: Energy deposition vs depth for various radiation modalities.

which beams are brought in from several angles and overlap at the tumor. Computer-controlled IMRT (Intensity-Modulated Radiation Therapy) adjusts the entry angle, dose per field and field shape (with multileaf collimators) to achieve excellent 3-D conformation for fields as large as 30 cm.

**Accelerators for x-ray therapy** 5-30 MeV S-band electron linacs are the mainstay of radiation therapy today (~5000 worldwide, principal manufacturers: Varian, Elekta, GE, Mitsubishi, Toshiba). X-rays are produced by electrons striking a heavy-metal target. The very broad bremsstrahlung spectrum is “hardened” by using absorbers to filter out contributions from lower energies. S-band linacs are a highly successful spin-off from high-energy and nuclear physics programs. Their compactness, efficiency and reliability have been key to their acceptance for clinical applications. With overall lengths of 1-2 m, and low rigidity of the electron beams, very compact accelerator and beam transport systems are possible. The development of the isocentric gantry (patient lies stationary while the x-ray beam is rotated around the tumor) facilitates IMRT. Powered by either a magnetron or a klystron, these accelerators operate at repetition rates up to 1 kHz.

**Proton and ion-beam therapy [7]-[13]** Fig.1 reveals the advantage of using heavy-charged particles (instead of  $e^-$ ) for therapy. First recognized by Bob Wilson in 1946, the Bragg Peak at the end of the particle’s range can provide a significant concentration of dose into a tumor. The wider proton curve arises from the higher multiple

scattering and range-straggling of protons, the tail of the carbon curve comes from nuclear breakup of the projectile into lighter (longer-ranged) fragments. Ionization density for charged particles varies as  $Z^2$ , so heavier ions in addition to having sharper stopping points are more lethal to malignant cells. Initial ion treatments used neon, but today carbon is the ion of choice for sparing of normal tissue on the entry path and good cell killing in the stopping region.

The well-defined stopping point also makes it easier to conform the radiation dose to an irregularly-shaped tumor by independently varying the position and energy of stopping particles. Pencil-beam scanning systems have been designed for this purpose, and are now in clinical use at several facilities.

Even with less sophisticated delivery systems, dose distributions of protons and ion beams are considerably better than even the best x-ray systems, and successful clinical programs with these beams have been ongoing for more than 50 years. Up until recently, the primary hurdle to widespread application has been limited availability due to its high cost.

**Accelerators for proton therapy** Proton energy of 250 MeV allows penetration to 30 cm in tissue. Average beam current of a few nA yields dose rates adequate for treatment times of ~1-2 min for all but the largest therapy fields. Early proton therapy was performed with fixed-energy synchrocyclotrons (e.g. Harvard, Uppsala); the large therapy fields ( $\approx 20$  cm dia.) at the required uniformity ( $\approx 5\%$ ) were produced using sophisticated scattering foil techniques; the range of the beam was varied by energy degraders in front of the patient. These “passive” delivery systems are decoupled from the accelerator, requiring only a fast and reliable beam cutoff system. In these first facilities, treatment port orientations were adjusted by moving the patient.

The more advanced delivery systems now being developed require energy variability, precise beam-intensity control and above all high stability and rapid response. The control systems for beam-delivery and accelerator must be tightly coupled. A continuous beam is preferred, with a duty factor  $> 25\%$ . Interfacing an advanced delivery system with a short-pulsed beam presents difficult problems.

The first hospital-based proton therapy accelerator was the synchrotron at Loma Linda,

## Sec.1.6: GLOSSARY OF ACCELERATOR TYPES

California, built by Fermilab. It has a weak-focusing lattice, injected by a 2 MeV RFQ with a single-turn kicker. Operating on a 2-s cycle, the half-integer resonant extraction provides reasonably flat spills with a 25% duty factor at any desired energy up to 250 MeV. Beam is transported to two fixed-beam rooms and three gantry rooms.

The field is advancing rapidly. For “conventional” layouts (one accelerator, and beam-lines serving several independent treatment rooms delivered as a “turnkey” system) IBA (Belgium) leads the pack with ten or more installations of their 235-MeV cyclotron-based system. Hitachi and Mitsubishi have between them eight slow-cycling synchrotron-based systems in operation. Accel GmbH developed a compact superconducting cyclotron (based on a Henry Blosser design) and has two operating facilities now. Accel’s cyclotron rights have been purchased by Varian that is now marketing this technology.

Energy changes in today’s facilities are mostly done outside the treatment room to avoid quality degradation and neutrons. Cyclotron-based systems use a degrader and collimated energy-selection channel in the transport line, adjustments take a few seconds. Ion source current in the cyclotron must be increased (by up to  $10^3$ ) to maintain brightness (dose rate). Synchrotron-based systems are capable of pulse-to-pulse energy variation.

Isocentric delivery is now standard. Because of the high rigidity of the proton beam (up to 2.5 T-m) gantries are very large (10-13 m dia). As much as 3 m between the last bending magnet and the patient (isocenter) is needed for treating large fields, and locating dosimetry and field definition equipment. Gantry diameter can be reduced by starting the spreading process before the last magnet, but this increases the size and weight of this magnet.

While passive (scattering-foil spreading) systems still account for the majority of installed delivery systems, active scanning is rapidly advancing. Next major technological hurdle is patient/organ motion during treatment, requiring online imaging instrumentation and live tracking by beam.

Developing compact accelerators for reducing size and cost of proton delivery is a hot topic today. Noteworthy is the Mevion 9T superconducting synchrocyclotron, mounted directly on a gantry in the treatment room. Delivery and in-

stallation of units started in 2012. The CPAC-LLNL high-gradient DWA (dielectric wall accelerator) induction linac aims for 100 MeV/m accelerating gradient, compact enough to also fit in the treatment room. Proton beams generated by high-power lasers are being developed, but are a long way from meeting clinical-beam specifications.

The “star dose” boost from capture of stopped  $\pi^-$  and  $\bar{p}$  excites interest in the physics community. Several thousand patients were treated with  $\pi^-$  in the 70’s/80’s at LAMPF, TRIUMF, PSI but disappointing clinical results halted trials.  $\bar{p}$  is out of economic reach for today’s technology.

FFAG’s (Sec.1.6.6) could provide energy variability with cyclotron-like beam quality, and rapid-cycling compact synchrotrons could provide cost savings. Linac energy-boosters are being developed in Italy.

**Accelerators for ion-beam therapy** An energy of 400 MeV/u is needed for a carbon beam to penetrate 30 cm tissue, corresponding to a magnetic rigidity of 6 T-m. Isocentric delivery presents a formidable challenge. Treatments with helium beams began in the mid 1950’s at Berkeley’s 184'' Synchrocyclotron, while trials with neon (and other ions) at the Bevalac took place between 1978 and 1993. All existing facilities are synchrotron-based (slow-cycling). Most now strive to deliver both proton and carbon beams from the same accelerator, for clinical intercomparisons.

The HIMAC facility in Chiba, Japan completed in 1994 has two 16 T-m synchrotrons (over/under) capable of 30 cm range with silicon beams and has treated 7000 patients with carbon. Hyogo, completed in 2001 has treated 4000 patients with carbon and protons. A third Japanese carbon facility at Gunma, is now running, and to more are under construction. GSI, starting in 1997 treated 440 patients with carbon, employing a sophisticated scanning system with control of all accelerator parameters for each pulse, from ion source out. The Italian CNAO facility in Pavia came online in 2011. All these offer fixed field orientation: either horizontal, vertical, or oblique for their carbon beams, but some do have gantries for proton delivery.

HIT in Heidelberg treated its first patient in 2010. This is a synchrotron-based facility with a 7 MeV/u RFQ/IH linac injection chain and two ECR-based ion-source front ends. It has two fixed-beam rooms, and one gantry room capable

of the full rigidity beam from the accelerator. The gantry is 13 m diameter, 25 m long, weights over 600 tons and has its scanning magnets upstream of the last 90° magnet.

Two new European projects: MedAUSTRON in Austria and ETOILE in France are under construction. At this time there are no active projects in the US.

Mitsubishi built the Gunma facility, and is actively pursuing further projects around the world.

IBA has designed a superconducting cyclotron capable of 400 MeV/u Q/A = 1/2 beams ( $C^{6+}$ ,  $H_2^+$ ), offering significant cost reduction and reduced facility size for ion beam therapy. Protons are extracted by stripping foil dissociation of the hydrogen molecule.

**Neutron therapy** Fast neutrons (14-70 MeV) have been used for therapy for over 50 years. With depth dose similar to lower-energy x-rays, localization of dose into a well-defined volume is difficult. However, the very high ionization density (owing to low-energy proton knockons from nuclear scattering) produces favorable results in radiosensitive tumors, located at shallow depths, particularly salivary-gland tumors. Interest in this modality is not widespread in the medical community. D-T generators (14 MeV neutrons) were widely used for many years. Few remain today because of maintenance issues and recognized need for higher-energy neutrons. Cyclotrons producing neutrons via either {p-Be} or {d-Be} reactions in the 60-70 MeV range were employed in the 1970-80. Of these, only the U. Washington machine remains. Also still operating today are the neutron therapy programs at Fermilab (60-MeV protons deflected onto a neutron target partway down the injector linac), and iThemba LABS, South Africa, using a 66 MeV extraction port from their 200 MeV separated sector cyclotron. A 60 MeV SC deuteron cyclotron mounted on a ring in the treatment room provides a compact source of neutrons from dissociation of the deuterons striking an internal target. This machine at Harper-Grace Hospital in Detroit is currently off line for economic reasons, but may restart in the future.

Slow neutron “capture therapy” (BNCT) has a small though faithful following. Boron-loaded, tumor-seeking pharmaceuticals are administered to the patient, the tumor area is flooded with epithermal neutrons which are selectively absorbed by the boron causing more radiation damage to the tumor tissue. Critical to success is the tumor-

specificity of the pharmaceutical. Though clinical success has been sparse, research continues. Reactor neutrons have been principally used in these studies. Initiatives now look to high-current (10-100 mA) low-energy (2.5 MeV) proton beams producing neutrons via the {p-Li} reaction. Also studied is driving subcritical reactor assemblies with high-current proton beams from FFAG structures.

### 1.6.14.2 Radioisotopes

Radioactive isotopes are widely used in both diagnostic and therapeutic applications [14]. Tracers, isotopes either alone or attached to physiologically relevant molecules, are used for functional imaging, detecting activity or hyperactivity concentrations in tissues. Imaging is possible using positron emitters (PET) such as  $^{11}C$ ,  $^{18}F$ , or single photon emitters and suitably collimated gamma-ray detectors (SPECT). PET isotopes are short-lived, and are produced with small single- or dual-particle (H or H/D) cyclotrons ( $\leq 18$  MeV) close to the end-use clinic.

While diagnostic isotopes emit hard gammas to minimize absorption in the body, alpha and beta emitters find therapeutic applications, maximizing dose to the volume close to the isotopic application, delivered either by uptake of injected/ingested material (e.g. iodine in thyroid treatments) or by surgical implantation of radioactive seeds (brachytherapy).

Accelerator-produced isotopes are made with proton (or alpha) beams of 35 MeV or lower from cyclotrons (e.g.  $^{201}Tl$ ,  $^{123}I$ ,  $^{67}Ga$ ,  $^{111}In$ ). Commercial production is concentrated in a few centers with elaborate distribution networks to provide rapid delivery of short-lived isotopes. Research isotopes are also produced at higher-energy accelerators [e.g. 100-800 MeV protons from TRIUMF, BLIP (AGS-BNL), LANSCE (LANL) and iThemba LABS (South Africa)].

$^{99}Mo/^{99m}Tc$  is the workhorse of nuclear medicine today, over 50,000 procedures are performed daily in the US. It is a prominent fission fragment from HEU (highly-enriched uranium) reactor cores. The supply of this isotope is not secure today owing to the age and reliability of the few production reactors, none of which are located on US soil. Interesting options exist for accelerator-production of this isotope using high-current SC linacs producing milliamperes beams of  $> 200$  MeV protons.

## Sec.1.6: GLOSSARY OF ACCELERATOR TYPES

**Accelerators for isotope production** PET isotopes are most commonly produced by (p,n) reactions with low-energy (11-15 MeV) cyclotron beams. The commercially-available cyclotrons are compact, self-shielded, highly reliable, and totally automated. Targetry and autochemistry units are usually included, providing complete hands-off preparation of isotopes in a form ready for administration. Manufacturers include CTI (Knoxville, TN), IBA (Belgium) and Ebcu (Vancouver).

Small cyclotron technology has been revolutionized by the development of high-quality H<sup>-</sup> ion sources, which solved the thermal, mechanical and activation problems associated with beam extraction. For energies up to 30 MeV, magnetic fields in the cyclotron can be high, leading to very compact structures, but for higher energies the magnetic field must be reduced to avoid Lorentz stripping (Sec.7.1.8) of the H<sup>-</sup> ions. TRIUMF (500 MeV) has a maximum field of 5 kG.

Beam currents for the PET isotope systems are modest (e.g. 50  $\mu$ A). The higher-energy cyclotrons used for production of longer-lived isotopes push the limits of current (up to 1 mA).

**Advanced radiography [15]** The advent of high fluxes of high quality x-rays from synchrotron radiation sources has opened up opportunities for diagnostics with monochromatic x-rays. A notable example is the coronary angiography program started at SSRL and continued at NSLS (BNL) and HASYLAB (DESY), in which an exposure taken with x-rays just above and just below the K edge are subtracted, producing a detailed image of coronary arteries with little contrast agent. Relevant x-ray energies are in the 10's of keV range. Storage rings e<sup>-</sup> or e<sup>+</sup> beam energies  $\gtrsim 2.5$  GeV are adequate for this purpose. For lower energy rings, high magnetic field dipoles and wigglers are required.

## References

- [1] R.R. Wilson, Radiology 47 (1946) 487
- [2] W. Scharf, Biomedical Particle Accelerators, AIP Press (1993)
- [3] W.H. Scharf, O.A. Chomicki, Physica Medica XII (1996) 199
- [4] Reviews of Accelerator Science and Technology, A. Chao, W. Chou, ed, Vol.2, World Scientific (2009)
- [5] C.J. Karzmark, C.S. Numan, E. Tanabe, Medical Electron Accelerators, McGraw-Hill (1993)
- [6] F.M. Khan, The Physics of Radiation Therapy, 2nd Ed., Williams & Wilkins (1994)
- [7] Particle Therapy Cooperative Group (PTCOG) (<http://ptcog.web.psi.ch/>)
- [8] Ion Beams in Tumor Therapy, U. Linz, ed, Chapman & Hall (1995); <http://juwel.fz-juelich.de:8080/dspace/handle/2128/659>
- [9] Ion Beam Therapy, U. Linz, ed, Springer, Biological and Medical Physics series (2012)
- [10] Hadrontherapy in Oncology, Proc. 1st Int. Symp. on Hadrontherapy (1993), Elsevier Excerpta Medica, Int. Congress Series 1077 (1994)
- [11] W.T. Chu, B.A. Ludewigt, T.R. Renner, RSI 64 (1992) 2055
- [12] W.T. Chu et al, LBL-33749 (1993); [http://www.osti.gov/bridge/product.biblio.jsp?query\\_id=0&page=0&osti\\_id=10163935](http://www.osti.gov/bridge/product.biblio.jsp?query_id=0&page=0&osti_id=10163935)
- [13] A.M. Sessler, convenor, Workshop on Hadron Beam Therapy of Cancer, Erice, 2009; <http://erice2009.na.infn.it/programme.htm>
- [14] NSAC\_Isotopes Subcommittee; <http://www.phy.anl.gov/mep/NSACI/>. Relevant reports: <http://www.sc.doe.gov/np/nsac.html>, “Compelling Research Opportunities using Isotopes” (Apr 2009); and “Isotopes for the Nation’s Future” (Nov 2009)
- [15] W. Thominson, NIM A319 (1992) 295

### 1.6.15 $\mu^+ \mu^-$ Collider *R. Palmer, BNL*

The idea of muon colliders was introduced by Skrinsky et al [1]. Neutrino Factories [2] use many of the same components.

#### 1.6.15.1 Collider

The main advantages of muons for a collider, as opposed to electrons, are:

- Synchrotron radiation,  $\propto E^4/m^4$ , requiring high energy electron colliders to be linear and long, is suppressed, so that a muon collider can be circular and smaller.
- Because circular, the muon bunches collide many times, allowing larger emittances for given luminosity. The number of such collisions is limited by the muon lifetime to  $n_{\text{turns}} \approx 150 B_{\text{ave}}$ , where  $B_{\text{ave}}$  is the average ring bending field in T.
- Synchrotron radiation, as bunches cross (beamstrahlung, Sec.2.5.3), is suppressed, resulting in much smaller collision energy spreads.

Table 1: Parameters of Collider Rings.

| C of m Energy TeV                                     | 1.5         | 3           |
|---|-------------|-------------|
| Luminosity $10^{34} \text{ cm}^2 \text{sec}^{-1}$     | 1.1         | 4           |
| Beam-beam Tune Shift                                  | .087        | .087        |
| Muons/bunch $10^{12}$                                 | 2           | 2           |
| Ring $\langle \text{bending field} \rangle \text{ T}$ | 5.8         | 8.4         |
| $\beta^*$ at IP = $\sigma_z$ mm                       | 10          | 5           |
| rms momentum spread %                                 | 0.1         | 0.1         |
| Repetition Rate Hz                                    | 15          | 12          |
| Proton Driver power MW                                | $\approx 4$ | $\approx 4$ |
| $\epsilon_{\perp}$ Trans Emittance $\mu\text{m}$      | 25          | 25          |
| $\epsilon_{\parallel}$ Long Emittance mm              | 72          | 72          |

- $S$ -channel Higgs production is enhanced by a factor of  $(m_\mu/m_e)^2 \approx 40000$ .

But there are challenges: selecting polarized muons is very inefficient; ring magnets and detectors must be shielded from decay electrons; acceleration and cooling must be fast to avoid decay losses; and, at higher energies, neutrino radiation is a significant hazard.

Representative parameters of colliders at 1.5 and 3 TeV CM energies are given in Tab.1. The initial muon 6-D emittance ( $\epsilon_{\perp}^2 \epsilon_{\parallel}$ ) is  $\sim 4 \times 10^6$  times the specified final value.

Synchrotron, electron, and stochastic cooling are too slow, but ionization cooling (Sec.2.7.4) appears feasible. Muons passing through an absorber, lose momentum in 3 directions, while only the longitudinal momentum is restored by rf. In the linear case, the equilibrium emittance in this process is

$$\epsilon_o \approx \frac{\beta_{\perp}}{\beta_v} \frac{1}{2} \left( \frac{14.1}{m_\mu} \right)^2 \frac{1}{L_R d\gamma/ds}$$

where  $\beta_{\perp}$  is the beam Courant-Snyder parameter,  $\beta_v = v/c$ ,  $m_\mu$  is the muon mass in MeV,  $L_R$  is the material radiation length,  $m_\mu d\gamma/ds$  is the energy loss per unit length. Hydrogen is the best absorber material. Efficient cooling requires  $\beta_{\perp}$  to be tapered to lower values as the emittance falls.

For 6-D cooling, the lattice must have dispersion and an absorber geometry that yields greater energy loss at higher momenta than lower. This reduces the longitudinal emittance, but increases the transverse, giving ‘emittance exchange’. Combined with transverse cooling this gives 6-D cooling.

The schematic of the most complete conceptual design [3] is shown in Fig.1. Fig.2 gives the longitudinal vs. the transverse emittances in the sequence of beam manipulations prior to muon

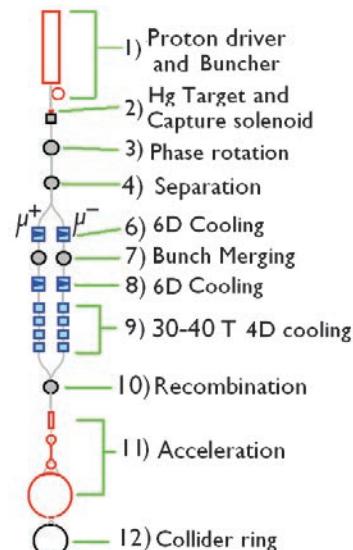


Figure 1: schematic of Muon Collider.

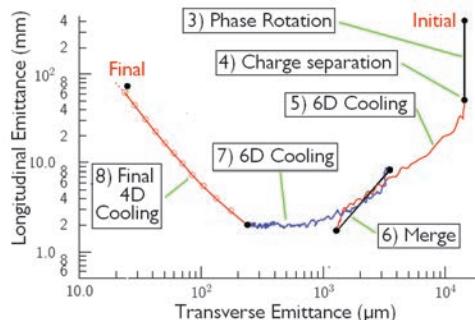


Figure 2: Longitudinal vs. transverse emittances in beam manipulations prior to acceleration.

acceleration. All components of this scheme have been simulated at some level. The main components, numbered as in Figs.1 & 2, are

1. A high intensity 8 GeV proton source, and a buncher that forms intense short ( $\sim 2$  ns) bursts.
2. A liquid metal target, able to withstand the 4 MW beam, in a 20 T hybrid solenoid (water cooled copper coils inside superconducting) to capture the pions. A tapered solenoid transports them to a lower field decay region.
3. A system of rf cavities bunch the muons and phase rotate [4] them into a train of both muon signs.

## Sec.1.6: GLOSSARY OF ACCELERATOR TYPES

4. Charge separation in a bent solenoid.
5. Separate 6-D cooling of each sign in periodic lattices following slow (Guggenheim) helices [7] (Fig.3b). Energy loss is in wedge shaped liquid hydrogen absorbers, focusing and bending by tilted solenoids, and re-acceleration in vacuum rf cavities. The Helical Cooling Channel (HCC) [8] (Fig.3c) is an alternative 6-D cooler. It is filled with high pressure hydrogen gas that acts as the absorber and suppresses rf breakdown.
6. The multiple bunches are merged into single bunches, one of each sign. Merging in longitudinal phase space is done by phase rotations using rf and drifts. Merging in the transverse dimensions is done by kicking different bunches into transports of differing lengths (a trombone[9]) to bring them at the same time to differing transverse positions.
7. 6-D re-cooling, of the now larger combined bunches, are continued to the lowest technically feasible transverse emittance. The transverse emittances at this point are still  $\sim 10$  times worse than those required, but the longitudinal emittances are  $\sim 100$  times better than needed.
8. Final 4D transverse cooling to the required transverse emittance is achieved with liquid hydrogen in high field (30-40 T) solenoids, and low energies, where the longitudinal emittance rises rapidly from the adverse dependence of energy loss on energy. But this is acceptable.
9. The muons of the two signs can now be re-combined, with suitable spacing.
10. Initial acceleration is in linacs, with frequencies first low, but rising as the bunches become shorter. By 200 MHz, the cavities should be superconducting. The linacs are followed by Recirculating Linear Accelerators, and then one or more pulsed synchrotrons [10].
11. The collider ring [11] must be isochronous to allow very short (1 cm) bunches. Chromaticity from the IP is locally corrected to allow low  $\beta^*$  insertions. Tungsten shielding is needed to keep the decay electrons from heating and quenching the ring's SC magnets. The detector also requires special shielding from these electrons.

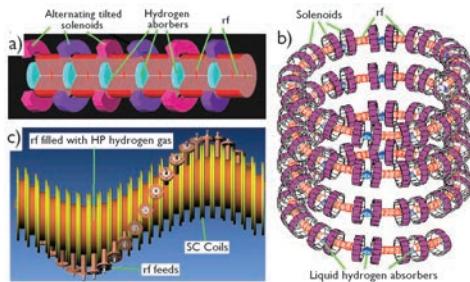


Figure 3: 6-D cooling lattices: a) HFOFO Snake; b) Guggenheim; c) HCC.

### 1.6.15.2 Muon storage ring neutrino factories

The main components of a neutrino factory [6] are the same as for a collider, but the requirements are less severe: more smaller muon bunches; far less cooling (a factor  $\sim 10$  vs.  $\sim 10^6$ ); no emittance exchange, and acceleration to lower energies (4-40 GeV vs. 0.75-1.5 TeV). Instead of a collider ring, the beam is injected into one or two long race track shaped storage ring, whose orientations and tilts are such that one of the long straight sections points to distant neutrino detectors.

### 1.6.15.3 Technical challenges

A liquid mercury target has been tested [12] in a proton beam at CERN and has demonstrated multi-megawatt capability.

An experiment to demonstrate ionization cooling [13] is under construction at Rutherford Appleton Lab (RAL) in the UK.

Experiments with rf in magnetic fields, at 805 [14] and 201 [15] MHz, have shown damage and/or limited acceleration gradients. It is proposed [18] that field emitted electrons are accelerated by the rf and focused by the magnetic fields, causing damage by cyclic heating. Several possible solutions are under study:

1. rf in high pressure hydrogen gas does not [16] show the problem, and the effects of a relatively low intensity proton beam did not cause breakdown [17]. With a small admixture of an electro-negative gas, the loss of rf power appeared acceptable [17].
2. Tests, in fields [14], showing severe damage of copper surfaces, have shown little or no damage on beryllium surfaces. A cavity with all beryllium end walls is under construction to test if this is a solution.

3. Improvements in surface treatments, including Atomic Layer Deposition [19] should suppress the initial electron production, and there are plans to test this.

Another challenge is to build the small bore 30-40 T ‘all superconducting’ solenoids for the final cooling. Lower fields give somewhat lower performance. HTS materials have sufficient current density for even higher fields, but it has not been demonstrated. A test that could reach 40 T is under construction [20].

## References

- [1] V.V. Parkhomchuk, A.N. Skrinsky, Proc. 12th Int. Conf. on High Energy Acc. (1983) 485; A.N. Skrinsky, V.V. Parkhomchuk, Sov. J. of Nucl. Phys. 12 (1981) 3
- [2] N. Holtkamp, D. Finley. eds., Fermilab-Pub-00/108E (2000)
- [3] R.B. Palmer, Muon Collider, Proc. AHIPA09
- [4] D. Neuffer, MUC-NOTE-269 (2003)
- [5] Y. Alexahin, <http://www.cap.bnl.gov/mumu/conf/collider-091201/talks/YAlexahin-1-091201.ppt>
- [6] J.S. Berg et al, PRST-AB 9 (2006) 011001
- [7] R.B. Palmer et al, PRST-AB 8 (2005) 061003; Pavel Snopok, <http://www.cap.bnl.gov/mumu/conf/MUTAC-090406/talks/PSnopok1-090406.pdf>
- [8] Y. Derbenev, R. P. Johnson, PRST-AB 8, 041002(2005)
- [9] C. Ankenbrandt, <http://indico.fnal.gov/getFile.py/access?contribId=65&sessionId=28&resId=0&materialId=slides&confId=2854>
- [10] D.J. Summers et al, PAC 07, THPMS082
- [11] Y. Alexahin, E. Gianfelice-Wendt, <http://www.cap.bnl.gov/mumu/conf/collider-091201/talks/YAlexahin-2-091201.ppt>
- [12] <http://proj-hiptarget.web.cern.ch/proj-hiptarget/>
- [13] MICE, <http://mice.iit.edu/>
- [14] A. Moretti et al, PRST-AB 8, 072001 (2005)
- [15] A. Bross, <http://www.fnal.gov/projects/muon-collider/FridayMeetings/> (2009)
- [16] P. Hanlet et al, EPAC 2006, TUPCH147
- [17] K. Yonehara, et al, MOPPC036, Proc. IPAC12, New Orleans, LA May 21-25, 2012
- [18] R.B. Palmer et al, PRST-AB 12, 031002 (2009)
- [19] S.M. George, A.W. Ott, J.W. Klaus, J. Phys. Chem. 100, 13121 (1996)
- [20] R. Weggel et al, NFMCC-DOC-553

### 1.6.16 Neutron Sources

*J. Wei, Michigan State U.  
H. Lengeler, CERN*

Neutron production methods include fission (reactor-based), isotope and excited-state decay, fusion (accelerated deuteron beam bombarding deuterium or tritium targets),  $e^-$ -bremsstrahlung-induced photon production and photonuclear reaction, charged-particle reaction, and spallation [1]. The latter three accelerator-driven sources rely on charged particle beams bombarding metal targets to produce n beams. Targets are surrounded by a moderator-reflector layout matched to produce neutrons in the desired energy range (cold, thermal, epithermal). Beam channels bring moderated neutrons to the experimental stations surrounding the shielded target station.

Because of the weak interaction of neutrons with matter and the engineering heat transfer limit on n production methods, n-scattering is an intensity limited field. In contrast to reactor sources, which are essentially c.w., accelerator-driven sources have the advantage that they can be pulsed. High peak flux is of interest for most n-scattering types of measurements. Accordingly, there are efforts in:

- Increasing the power of the incident beam, realized by increasing the beam energy, beam intensity, and using a pulsed time structure.
- Increasing the n target yield. Neutron yield increases significantly with the incident-particle energy, and is a strong function of the target material and incident beam type.
- Increasing the n transmission efficiency. The target-moderator-reflector design is optimized; guides with high-reflectivity, sometimes with focusing properties, are used.

According to the incident beam’s time structure, there are three types of accelerator-driven sources (Tab.2)[6, 7, 8]:

- C.w. sources driven by high-intensity electrostatic accelerators or cyclotrons. Example: the isochronous-cyclotron of PSI, which delivers a c.w. proton beam of 1.3 MW at 590 MeV [9].
- Sources with long (ms) pulses driven by high-intensity linacs. Examples: the 800

## Sec.1.6: GLOSSARY OF ACCELERATOR TYPES

MeV LAMPF p-linac at Los Alamos [5], the 13 MeV LENS p-linac at Indiana University, and the proposed 2.5 GeV ESS p-linac in Lund.

- Sources with short ( $\mu$ s) pulses driven by a combination of high-intensity linacs and rings. Example: the J-PARC 3 GeV proton rapid-cycling-synchrotron (RCS) supplied by the 181 MeV  $H^-$  linac and the SNS 1 GeV proton accumulator ring (AR) supplied by the 1 GeV  $H^-$  linac.

A dominant design criterion is ultra low beam losses for avoiding component, air, and tunnel activation and for allowing maintenance and repair on short notice. A typical guideline is to limit the average uncontrolled beam loss to be below 1 W/m.

**Compact neutron sources** In fusion-based (D-D, D-T, T-T) n generators, beams of currents up to 350 mA are accelerated with a voltage typically from 60 to 120 kV. The n yield covers a wide range from  $10^6$  to  $10^{11}$  n/s. Compact sources based on electron and proton linacs produce yields up to  $10^{14}$  n/s (Tab.1).

**Spallation neutron sources** In spallation sources [5], neutrons are produced by the interaction of high energy (GeV) protons with a target. The number of n produced is proportional to the proton energy in range 0.2-10 GeV; in a typical lead target a 1 GeV proton produces 20 neutrons.

Pulsed sources allow time-of-flight correlation of the incident neutron energy. This avoids the monochromatization needed in c.w. sources which greatly reduces the useful n-flux. Time of flight measurements require small repetition rate ( $\leq 60$  Hz), in order to avoid the overlap of slow n from one pulse with fast n from the next. Long-pulse (ms) facilities have the advantage of a simpler driver layout (i.e., a pulsed linac with high proton intensity), coupled target-moderator configuration, and potentially higher integrated n-flux. Short-pulse ( $\mu$ s) facilities have the advantage of much higher peak n-flux. However, they require an  $H^-$  linac combined with an AR or RCS filled by multитurn injection and emptied by fast one-turn extraction (Fig.1). Existing and planned spallation sources use a proton energy range 0.5-5 GeV. Different options for linac and ring energies have been used or proposed (Tab.2) [5, 6, 8], depending on the goal beam power.

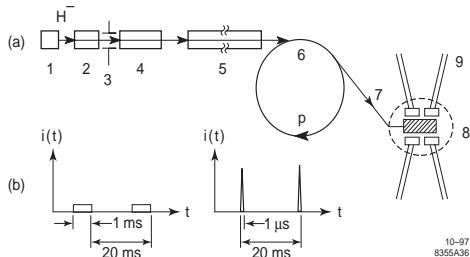


Figure 1: (a) Schematic of a short-pulsed neutron source. 1)  $H^-$  source, 2) RFQ, 3) chopper, 4) DTL or CCDTL, 5) nc or sc high energy linac, 6) AR or RCS with  $H^-$ - $H^+$  charge exchange injection, 7) beam transport to target, 8) target with moderators and reflectors, 9) neutron channels. (b) Time structure of beams at linac and at ring.

**Linac** Linacs for long-pulse sources accelerate high-intensity protons; up to 200 mA peak current is at the limit of present-day technology. Linacs for short-pulse sources accelerate high-intensity, low-emittance  $H^-$  for low-loss charge-exchange injection into the ring; up to 100 mA peak current and down to  $\sim 0.1 \pi$  mm-mrad normalized rms emittance are at the limit of present-day technology. Duty cycles of up to 100% has been demonstrated for RFQ and superconducting rf linac (SCL).

A linac starts with a proton or  $H^-$  source (Sec.7.1.4) followed by one or more sections of SC and/or non-SC accelerating or transport structures. The 1-GeV SNS linac consists of  $H^-$  source, LEBT, RFQ (Sec.1.6.18), MEBT, DTL (Sec.7.3.5), and SCLs of two different  $\beta$ 's. The 400-MeV J-PARC linac consists of  $H^-$  source, LEBT, RFQ, MEBT, DTL, SDTL, and ACS.

SCL is preferred to non-SC options especially for duty cycles higher than 5%, and for retuning and operational robustness. For a pulsed p beam with its velocity changing during acceleration, effects of Lorentz detuning, microphonics, beam transients and injecting-energy offsets require careful rf amplitude and phase control [2]. The SNS's SCL (6% duty) starts at 186 MeV with each cavity driven by its own klystron and rf control.

To keep average linac losses low ( $\sim 1$  W/m) emittance growth and halo generation have to be carefully controlled at all stages of the linac, in particular at structural and rf frequency transitions. Low loss ring injection requires new ways

Table 1: A few existing and planned compact sources.

| Name           | Status  | Accelerator type & type energy | Target | Ave. beam power (kW) | Rep. rate (Hz) | Pulse length at target (ms) | Ave. n-yield ( $10^{13}$ ) (n/s) |
|----------------|---------|--------------------------------|--------|----------------------|----------------|-----------------------------|----------------------------------|
| Hokkaido       | operat. | 45 MeV e-linac                 | W+Pb   | 1                    | $50-100$       | $10^{-8}-3 \times 10^{-3}$  | 0.16                             |
| LENS, Indiana  | operat. | 13 MeV p-linac                 | Be     | 13                   | 20             | 2                           | 4                                |
| CPHS, Tsinghua | constr. | 13 MeV p-linac                 | Be     | 16                   | 50             | 0.5                         | 5                                |

Table 2: A few existing and planned spallation sources.

| Name               | Status                 | Accelerator type & energy       | Target type | Average beam power (MW) | Rep. rate (Hz) | Protons per pulse ( $10^{13}$ ) | Pulse length at target ( $\mu$ s) |
|--------------------|------------------------|---------------------------------|-------------|-------------------------|----------------|---------------------------------|-----------------------------------|
| IPNS, ANL [5]      | 1981–2008              | 50 MeV linac<br>500 MeV RCS     | U           | 0.0075                  | 30             | 0.3                             | 0.1                               |
| ISIS, RAL [10]     | operat.<br>1985–       | 70 MeV linac<br>800 MeV RCS     | Ta, W,<br>U | 0.16                    | 50             | 2.5                             | 0.45                              |
| SINQ, PSI [9]      | operat.<br>1996–       | 590 MeV cyclotron               | Zr,<br>Pb   | 1                       | cw             | -                               | -                                 |
| LANSCE<br>LANL [5] | 1985–PSR<br>1975–linac | 800 MeV linac<br>AR             | W           | 0.08                    | 20             | 3                               | 0.27                              |
| SNS, ORNL [11, 12] | operat.                | 1 GeV linac (nc/sc)<br>AR       | Hg          | 1                       | 60             | 10                              | 0.7                               |
| J-PARC [13] Tokai  | operat.                | 181(400) MeV linac<br>3 GeV RCS | Hg          | 0.3 (1)                 | 25             | 8.3                             | 1                                 |
| CSNS [16] Dongguan | constr.                | 80 MeV linac<br>1.6 GeV RCS     | W           | 0.1                     | 25             | 1.6                             | 0.8                               |
| ESS Lund           | planned                | 2.5 GeV linac                   |             | 5                       | 20             | 62                              | 2000                              |

of halo containment, diagnostics, collimation and beam cleaning. Extensive space-charge analysis of resonance and equipartition conditions and comprehensive Monte-Carlo simulations (including nonlinear space charge forces) have to be applied with up to  $10^7$  macro-particles for a realistic layout between  $H^-$ -source and ring injection [17]. Intrabeam stripping (Sec.2.4.12) is identified to be a loss mechanism for  $H^-$  beams and studied at the SNS linac [18].

The ultra low injection losses also require the transport (HEBT) between linac and rings to have a precise control of energy (energy-deviation correction by a rotator cavity and energy-width spreading by ramping or a spreader cavity) and a removal of halo particles by betatron and momentum scraping. At high injection energies, low magnetic fields have to be used to avoid magnetic stripping (Lorentz stripping, Sec.7.1.8) of  $H^-$ s.

At ultra-high energies (e.g. 8-GeV linac proposed by Fermilab), black-body stripping further limits the maximum field strength of the magnets [3].

**Rings** [19] The SNS AR of 248-m circumference accumulates 1 MW p-beam ( $2.9 \times 10^{-6}$  duty, 35 A peak current). The J-PARC RCS presently accelerates 300 kW p-beam from 181 MeV to 3 GeV. The average uncontrolled beam loss is  $\sim 1$  W/m limiting the performance. At very high beam power (e.g.  $> 5$  MW) the use of multiple rings may become mandatory.

Upon charge-exchange injection using a stripping foil, elaborate painting with correlated 4-D or 6-D phase space fills the ring acceptance as uniformly as possible. Large ring acceptance (typically  $> 400 \pi$  mm-mrad) and good magnetic field quality ( $\sim 10^{-4}$  deviation level) are needed for keeping effects of space charge and magnetic resonances at a tolerable level. Stripping foils have

## Sec.1.6: GLOSSARY OF ACCELERATOR TYPES

an efficiency of about 98-99% so that a high intensity of partially stripped H<sup>0</sup> particles has to be handled in the injection region. This low emittance beam presents an interesting source of high energy protons which can be used e.g. for the production of radioactive beams or muon beams.

The ring lattice is generally based on a high periodicity and transition energy (Sec.4.9 should not be passed). Long straight sections are used for rf, injection, extraction and scraping systems. The efficiency of rf trapping can be increased by using a low frequency, dual-harmonic rf-system and by chopping the injected H<sup>-</sup> beam (at the linac front end) at the ring revolution frequency.

Collective effects and instabilities are most relevant; low impedance beam lines are essential. For RCSs with ceramic vacuum chambers either internal rf-cages or external metal stripes are used [10, 13]. The coupling impedance of extraction kickers and the pulse-forming network is minimized [21, 8]. Effects of any beam-induced electron cloud are mitigated by vacuum-chamber surface coating, solenoidal magnetic fields, and clearing electrodes to reduce the secondary emission yield [22, 23, 8] (Sec.2.4.14). At ISIS, the tunes are adjusted to (i) compensate for the natural chromaticity and the varying magnet field at injection, (ii) ramp up tunes to minimize effects of space-charge depressions during beam capture, (iii) reduce tunes during the time from 2 to 4 ms after injection to avoid transverse resistive wall instability, and (iv) lower tunes to avoid coupling resonances at extraction.

Fast kicker magnets (< 200ns) (Sec.7.2.7) are needed to avoid beam losses at extraction. Particles are brought to the target station with large acceptance transfer lines which may contain multipole elements to flatten the beam density profile at the target.

Momentum scraping is essential for RCSs to control the beam loss during rf trapping and ramping. Betatron scraping efficiency is improved by a two-stage system with acceptances significantly smaller than the rest of the ring, but larger than the beam core [24]. Beam-in-gap cleaning reduces the beam loss at extraction for ARs [25, 24].

Adequate shielding, beam dumps and fast beam loss monitoring acting bunch-to-bunch are essential. Radiation resistant materials are used in areas of high radio-activation. Remote handling devices are used in areas demanding frequent maintenance. Large spallation sources have to

supply neutrons to a large number of experiments (~1000-2000/yr). High availability, high reliability and short down-times are essential.

**Neutron targets** [5] For spallation sources, water-cooled solid (Ta, W, Pb, depleted U) and liquid (Hg, Pb-Bi) heavy metal targets with horizontal or vertical injection are in use up to MW range. Rotating solid targets are studied for multi-MW class sources.

Targets have to contain the nuclear cascade produced by protons and withstand high radiation damage. For MW targets this can become comparable to the range expected in Tokamak fusion reactors (> 100 d.p.a. – displacements per atom – produced by p and n knock-on, and 10,000 appm – atomic parts per million – of He produced by transmutation).

For short pulses an additional load stems from shock/stress waves produced by the high energy content (up to 100 kJ) of proton pulses. This combined load and the absence of corrosion and tritium production linked to cooling water circuits are the main reasons for developing liquid targets. Liquid targets in pulsed operations suffer cavitation-induced pitting damage to the surfaces of target vessels. Stainless steel specially treated and hardened by the kolsterizing process is in use.

For compact sources driven by lower energy proton beams, water-cooled Li and Be targets are in use, and Ga-liquid-metal cooling has been proven out [26]. The short stopping range of low-energy p beams causes complications.

**Other applications** The accelerator and target technology necessary for spallation sources in the MW-range has requirements which are in many respects similar to the ones needed for future high power proton accelerators envisaged for neutrino factories, muon storage rings (Sec.1.6.15), transmutation of nuclear waste, accelerator production of tritium, accelerator driven subcritical nuclear power generation, proton radiography, white neutron source applications, and high intensity radioactive beams. Its use for irradiation facilities is also being considered.

## References

- [1] J.M. Carpenter, Y.B. Yelon, Methods of Experimental Physics, Ch.2, Vol.23, p.99 (Academic Press)
- [2] J. Wei et al, PAC 2001, 319
- [3] H.C. Bryant, G.H. Herling, J. Mod. Optics 53 (2006) 45

- [4] T.A. Gabriel, J.R. Haines, T.J. McManamy, *J. Nucl. Materials* 318 (2003) 1
- [5] Proc. Int. Collaboration on Advances Spallation-Sources (ICANS); ICANS XII, Abingdon, UK (1993), RAL Proc.94-025, ICANS XIII (1995) PSI Proc. 95-02
- [6] G.H. Rees, PAC 93, p.731
- [7] P. Bryant, PAC 95, p.322
- [8] J. Wei, RMP 75 (2003) 1383
- [9] G.S. Bauer et al, PAC 97
- [10] D.J. Adams et al, EPAC 00, 975
- [11] J. Wei et al, PRST-AB, 3 (2000) 080101
- [12] N. Holtkamp, Linac 04, p.837
- [13] JAERI/KEK Report JAERI-Tech 2003-044, KEK Report 2002-13 (2003)
- [14] ESS Study Final Report, Vol.III, ESS-96-53M (1996)
- [15] H. Lengeler, NIM B139 (1998) 82
- [16] IHEP Report IHEP-CSNS-Report/2004-01E (2004)
- [17] K. Bongerdt, M. Pabstand, A. Letchford, NIM A451/1 (2000) 287
- [18] V. Lebedev et al, Linac 2010, p.929
- [19] G.H. Rees, EPAC 94, p.241
- [20] R. Macek et al, PAC 01, p.688
- [21] D. Davino et al, EPAC 02, p.1467
- [22] P. He et al, EPAC 04, p.1804
- [23] L. Wang et al, PRE 036501 (2004)
- [24] N. Catalan-Lasheras et al, PRST-AB, 4, 010101 (2001)
- [25] R.L. Witkover et al, PAC 99, p.2250
- [26] B.W. Blackburn, MIT Thesis (2002)

### 1.6.17 Pulsed High Voltage Devices

*J.A. Nation, D.A. Hammer, Cornell U.*

The engineering science of Pulsed High Voltage devices [1, 2] was first developed by J.C. Martin of AWRE (Atomic Weapons Research Establishment, Aldermaston, UK). He used a Marx generator to impulse charge a solid dielectric, or oil-filled, or water dielectric transmission line as a lumped parameter capacitor. The discharge of the transmission line, now as a distributed line, into a vacuum diode reduced the pulse duration, compared to the charging time of  $\sim 1 \mu\text{s}$ , by an order of magnitude and hence increased the available power correspondingly. Electron beams were produced by field emission cathodes. Simple systems of this type produce output pulses of 1-10 MV with pulse durations of order 20-100 ns, at impedances of a few to 50  $\Omega$ , depending on the dielectric used in the pulse line. More recent extensions of the technology produced  $\sim 30$  MV

output voltage, sub-1  $\Omega$  impedances, and power levels of tens of TW. Applications of this technology include x-ray generation and inertial confinement fusion.

**Marx generators/Pulse lines** The typical Marx Generator uses plus/minus charged columns of capacitors that are charged in parallel and discharged in series, through triggered spark gaps, into a transmission line. The charging and discharging of the transmission line occurs in  $\sim 1 \mu\text{s}$  and  $\leq 0.1 \mu\text{s}$  respectively. The transmission lines use deionized water (for low impedance) or transformer oil (for intermediate-to-high impedance) as the insulating material. Breakdown strengths of the dielectrics are  $\sim 100$  and 300 kV/cm respectively for pulse durations  $\sim 1 \mu\text{s}$  and increase slowly ( $\propto t^{-1/3}$  for sub- $\mu\text{s}$  pulses) with decreasing pulse durations. The transmission line is usually connected to the load by an overvolted gas or water spark gap. Modest changes in the pulse duration and the generator output impedance are achieved through the use of tapered transmission lines connecting the pulse line to the load. In sub-1  $\Omega$  impedance generators the basic Marx generator-pulse line configuration may be repeated many times (e.g., the Z-machine at Sandia uses 36 modules in parallel). Very short duration high power pulses require low inductance power feeds, which is accomplished by the the use of water-dielectric transfer capacitors as an intermediate low inductance circuit element between the Marx generator and the transmission line. The latter can be charged with a faster rise time from the water capacitor than directly from the Marx generator, leading to a more compact system as well as to the formation of multiple channel discharges, and hence low series impedance, connecting the transmission line to the load. In ultra high current devices the load is connected to the transmission line by a Magnetically Insulated Transmission Line (MITL), an evacuated transmission line in which the wave electric field causes electron emission from the negative line conductor. However, the current in the MITL produces a sufficiently large transverse magnetic field between the line conductors to cause the emitted electrons to flow parallel to the electrodes instead of across the electrode gap.

**Transmission line loads** The load depends on the application. Three commonly used loads are: (i) Vacuum diodes with field emission cathodes that produce electron beams; (ii) Diodes with

plasma electrodes that are used to produce electron or ion beams; and (iii) Z pinches.

The above arrangements are commonly used for hard x-ray production, high power microwave generation, inertial confinement fusion research and for the generation of intense soft x-ray pulses. Recent experiments at Sandia have reported the production of 2 MJ x-ray pulses with peak powers of up to 280 TW.

Diodes naturally generate electron beam currents when a suitable polarity high voltage pulse is applied across the diode. Ion beams can also be generated if there is a suitable ion source, such as a plasma, on the anode, but the ion generation efficiency is low unless the electron current is suppressed. This is accomplished by applying a transverse magnetic field in the diode such that the electron excursion from the cathode is less than the anode cathode gap spacing. Proton and other low atomic number ion beams have been produced this way.

**Inductive addition** Induction accelerators are discussed in Sec.1.6.7. High voltage pulses for radiography or for use in electron beam injectors are frequently produced by the use of a single central cathode conductor as the secondary of several induction modules. The voltage of the cathode is then equal to the sum of the secondary voltage outputs of each of the modules; i.e. the conductor adds the voltages from the modules. For hard x-ray production, which scales as  $I \cdot V^{2.8}$  for a given high- $Z$  target material, the high voltage output is essential for efficient use of the accelerator power. The Hermes III accelerator at Sandia, e.g., uses 20 1-MV induction modules to produce a 20 MeV electron beam in a single diode. For electron beam production the increased injector energy allows larger space charge limited beam currents, and reduced beam divergence.

**Linear Transformer Driver (LTD)** The fundamental unit of LTD technology is a capacitor switched into a low inductance circuit with soft iron core isolation. In order to achieve a very high current in a low inductance load, such as a Z-pinch, many such modules are switched in parallel in a circular array using a large, circular iron core to isolate all of them, with the load in the center. Several 1 MA pulsers with rise times of 150-300 ns have been built using this approach [3]. Such circular arrays can then be stacked in series in order to drive a higher impedance load, or in series-parallel arrangements to reach ultra-

high power, as has been done in the design for the next larger Z-pinch driver by Stygar *et al.* [4].

**Voltage and current fluctuations** For many pulsed power applications, square voltage and current waveforms are of secondary importance compared to the peak power output. This is especially true in multi-TW devices where the low impedance of the generator leads to a relatively slow increase in the load current. For example, the Z-machine delivers a  $\sim$ 5 MV, 26 MA, 100 ns current rise-time pulse to a Z-pinch. The voltage fluctuations can be  $\sim$  50% and the current rise time is comparable to the pulse half width. The device is well matched, however, to the dynamics of the imploding Z-pinch. The fluctuations in the output of higher impedance devices are, however, much lower and a degree of tuning is possible, especially in the beam production mode of operation.

**Pulse transformers** Many of the above devices are not well suited to high repetition rate operation, except in burst modes. For applications requiring more modest beam currents, (e.g. klystrons), pulse transformers offer the preferred modulator configuration. Thyratron switching permits high repetition rate use more readily than that achievable with pressurized gas switches. Pulse durations are typically  $\sim$ 1  $\mu$ s and output voltages in the range 300-500 kV. The achievable rise time is strongly affected by the transformer step-up ratio and the core material selection. Rise times of 100-200 ns are achievable with step-up ratio's of  $\lesssim$  8.

## References

- [1] J.C. Martin, Pulsed Power, Advances in Pulse Power Vol.3, Plenum Press (1996)
- [2] J.A. Nation, PA 10 (1970) 1
- [3] A.A. Kim *et al*, PRST-AB 12, 050402 (2009)
- [4] William Stygar *et al*, PRST-AB 10, 030401 (2007)

### 1.6.18 Radio Frequency Quadrupole *J. Staples, LBNL*

A Radio Frequency Quadrupole (RFQ) is a compact and versatile accelerator operating over a mass range of protons to low charge state heavy ions, accelerating from a few keV/n for heavy ions with a total voltage integral of up to several MV. The operating frequency ranges from 6 MHz (for  $\text{Bi}^{+2}$ ) to over 400 MHz (for protons). Duty

factors range from 0.1% to 100%, and transmitted current to greater than 200 mA of protons.

**Beam dynamics** An RFQ comprises an alternating-gradient time-varying electric quadrupole strong focusing transport channel with an accelerating field  $E_z(z)$  added as a perturbation by *modulating* the vanetip profile. The  $E_z(z)$  profile may be chosen arbitrarily, giving wide design freedom of the capture and acceleration sections. The alternating-gradient electric-quadrupole field provides velocity-independent focusing and will transport unaccelerated or partially accelerated beam to the exit.

As a *Sloan-Lawrence* accelerator, the shunt impedance drops off as  $\beta^{-2}$  in the accelerating section, restricting output beam energies to less than 2-3 MeV for protons for efficient designs.

$E_z(z)$  is controlled by the vane modulation parameter  $m \geq 1$ , where the displacement of the vanetip from the axis varies from  $a$  to  $ma$  along the length of the cell.  $E_z(z)$  scales approximately as  $(m-1)^{2/3}$ . The upper limit of  $m$  is constrained by the *minimum longitudinal radius* which sets the size of the tool used to cut the vane profile, as well as the magnitude of the multipole components of the fields. As  $m$  and thereby  $E_z$  increase, the transverse focusing strength and the clear beam aperture decrease, limiting geometric acceptance.

The transverse phase space acceptance of an RFQ increases rapidly with the design field gradient. The peak surface field on the vanetip is  $E_s = \kappa V/r_0$ , where  $\kappa$  is the field enhancement factor.  $V$  is the peak rf voltage between vanetips, and  $r_0$  is the average vanetip displacement from the beam axis. Depending on the detailed vanetip geometry,  $\kappa \approx 1.25 - 1.55$ . The sustainable field is expressed in units of kilpatrick, found by solving the implicit equation  $f = 1.643E_s^2 e^{-(8.5/E_s)}$  for  $E_s$ , the surface field in MV/m,  $f$  is the frequency in MHz. One kilpatrick at 200 MHz is 14.7 MV/m, and 19.4 MV/m at 400 MHz. Short-pulse RFQs may be safely pushed to greater than 2.0 kilpatrick, with c.w. RFQs held to less than 1.8 kilpatrick.

**Beam dynamics codes** RFQ design and simulation codes include PARMTEQM [1], TOUTATIS [2], LIDOS [3] and RFQtrak [4]. PARMTEQM uses the *Kapchinskii-Teplyakov* (KT) design procedure where the beam is bunched adiabatically with a long *gentle buncher* section, typically oc-

cupying more than half of the physical length of the RFQ, followed by an *accelerating* section. The d.c. input beam is matched at the RFQ entrance to the time-varying transverse beam profile by the *radial matcher*, typically 4-8 cells, followed by a short *shaper* section that initiates the formation of the bunch. The helper code RFQUIK assembles a cell table for use in PARMTEQM. To reduce the truncation error of the two-term field expansion, PARI adds higher-order terms and adjusts the cell parameters to maintain the acceleration of the reference particle. PARI uses a look-up table for a limited number of vanetip profiles.

The TOUTATIS code, requiring significantly more computational resources, solves the boundary-value problem for the actual vanetip geometry in each cell, allowing arbitrary cell shapes and gaps in the vane, and a PIC formulation of the space charge forces. TOUTATIS is based on PARMTEQM, but removes the paraxial approximation, adds field maps, and treats transverse beam loss on the physical location of the vane boundary. The LIDOS and RFQTRAK codes also include field maps for a more accurate evaluation of the beam characteristics. Jameson [5] presents a lengthy summary comparing the above codes, used during the design of the IFMIF c.w. RFQ.

The KT formalism is suitable for the design of high-current RFQs, but greater bunching efficiency, lower emittance growth and shorter structures may be generated using other design approaches. None of these approaches is straightforward, but rather ad-hoc, without a specific design recipe. The Hofmann diagram [6], which illustrates bunch resonances as a function of betatron phase advance and tune depression, is of use in determining an optimum parameter space for high-current, low emittance-growth designs.

**Space charge** The KT approach used in PARMTEQM uses the helper code CURLI to optimize the beam dynamics design at the end of the gentle buncher section where the bunch is formed, the charge density is high and the energy is still low, so the tune depression is at a maximum. The transverse and longitudinal current limits  $I_t$  and  $I_\ell$  are defined as the current that depresses the tune by typically 60%. The transverse limit  $I_t$  scales as  $\beta\lambda^2 V^2 \phi_x q / (a^2 A)$  and the longitudinal  $I_\ell$  as  $V\phi_s^2 a / \lambda$ , with  $a$  the minimum vane tip radius,  $V$  the vane-to-vane peak voltage,  $\phi_s$  the stable phase, and  $q/A$  the charge-to-mass of the ion, indicating that a low-frequency machine is

preferred for a large transverse current limit, but a high-frequency machine would have a higher longitudinal current limit. Maintaining an adequate focusing phase advance for mass greater than proton requires a longer operating wavelength  $\lambda$ , as the vane voltage  $V$  and aperture  $a$  are already at a practical limit. The operating frequency of heavy-ion RFQs may be in the 50-200 MHz range, or even lower.

**Rf structures** All RFQs have in common a time-varying electric quadrupole focusing field on axis. The choice of rf structure will depend on the operating frequency, length, and power efficiency required.

The *4-vane* structure excites a waveguide in the TE<sub>210</sub> mode, producing a quadrupole E-field, adding vanes to concentrate the electric field near the axis. A typical frequency range is 100 to 425 MHz. A major drawback of the 4-vane structure is the presence of almost degenerate dipole TE<sub>110</sub> modes which may mix with the TE<sub>210</sub> quadrupole mode. Several methods have been used to increase the dipole-quadrupole frequency separation. Opposite vanes may be strapped together with vane coupling rings (VCRs) in low duty-factor structures. or  $\pi$ -mode stabilizers for high duty-factor RFQs, used at JPARC and SNS. The LEDA RFQ uses resonant coupling between longitudinal RFQ regions to effectively shorten each section, reducing the field perturbations due to local frequency error.

The *4-rod* structure avoids the TE<sub>110</sub> mixing problem by supporting opposing vane pairs on quarter-wave stubs, shortened due to the additional capacitive loading of the vanes themselves. The electric field energy is almost entirely within the vane region, and the magnetic field energy mainly around the stubs. The field profile may be modulated by the periodic structure of the support stubs, and there may be a non-zero potential between the ends of the vanes and the end-walls, affecting bunching. Both of these potential problems are manageable. This structure is more compact than the 4-vane structure and has been frequently applied to lower-frequency heavy-ion accelerators.

The *split-coaxial* structure has been promoted for even lower frequency accelerators for low-charge-state heavy ions and for superconducting RFQ structures. Here, the 4-vane structure is modified with large cut-outs in the vane base that modify the mode structure, moving the TE<sub>110</sub>

dipole modes higher in frequency and away from the quadrupole mode.

**Rf structure codes** The 4-vane RFQ operates at the waveguide cutoff frequency, which may be estimated with 2-D codes such as SUPERFISH [1] or URMEL. However, the presence of mode stabilizers and vane end cutbacks require a full 3-D calculation for accurate estimate of the resonant frequency, field distribution and wall power density.

Advances in computational capabilities have resulted in the development of powerful finite-element and finite-difference codes that permit the characteristics of complex rf structures to be determined. RFQs, particularly ones that employ various types of stabilizers and tuners, require a large range of mesh density to include small details over a large cavity volume.

Example codes include CST-Microwave Studio [9], HFSS [10], and Omega3P [11], which have been used to model RFQ mode structure and surface power density. CAD modeling codes such as ANSYS [12], which also include 3-D electromagnetic solvers as an additional module are used to model the time-dependent thermal and thermally-induced stresses in the structure.

**Tuning** The 4-vane RFQ is a standing-wave structure where the deviation of local field  $dE_0(z)/E_0$  as a function of the local detuning  $\delta f_0(z)/f_{\text{average}}$  is given by the solution of

$$\frac{\partial^2}{\partial z^2} \left( \frac{\delta E_0(z)}{E_0} \right) = \frac{8\pi^2}{\lambda^2} \left( \frac{\delta f_0(z)}{f_{\text{average}}} \right) \quad (1)$$

where  $\delta f_0(z)$  is the local frequency variation due to mechanical errors and  $\int(\delta f_0) = 0$ . Field errors scale as the square of the length of the RFQ. This restricts the practical length of an RFQ to less than about 5 free-space wavelengths above which mechanical tolerances become severe.

The field distribution is usually determined by pulling a metallic or dielectric bead through the RFQ and noting the change in resonant frequency. The metallic bead removes both E- and H-field, the dielectric bead only E-field stored energy. Sensing loops may be placed along the RFQ and calibrated by the bead pull to measure the fields during operation.

Tuners are introduced along the outer walls of 4-vane RFQs usually in the form of pistons which when moved inward, increase the local resonant frequency by removing H-field energy. A program that specifies tuner settings by solving

Eq.(1), based on bead pull measurements, eases tuner adjustment.

The field distribution in 4-rod RFQs concentrates the magnetic field energy in the volume around the support studs. Tuners may be in the form of capacitive plates near the rods, lowering the local frequency, or in metallic blocks in the vicinity of the support studs, raising the local frequency.

**Rf power couplers** The rf may be introduced through multiple or single loop or iris couplers, which couple with the rf H-field, maximum at the wall of 4-vane RFQs. Loop couplers are easily adjustable, but are limited in the peak rf power that can be introduced in each coupler to the few hundred kW range. Multiple couplers driven from the same rf source should be isolated from each other to ease balancing the power flow through each.

Higher average power may require the use of iris couplers. The high power density at the edge of the iris will require special attention.

The measured Q-value of typical 4-vane structures runs from 50 to 80% of the theoretical value calculated by the electromagnetics codes.

**High average power RFQs** RFQs designed for c.w. operation include LEDA [8], TRASCO [13], IFMIF [14], Project-X(PXIE) [7] and many others. Pisent [15] describes several specific designs. Designs tend to lower frequency to reduce the wall power density to keep it below the 15 W/cm<sup>2</sup> range with larger aperture to reduce beam loss in the structure. High duty-factor RFQs with high average wall power densities (in excess of 1 W/cm<sup>2</sup>) may change shape due to thermal expansion, resulting in a redistribution of the field profile. At 400 MHz, an isothermal copper cavity shifts 7 kHz per degree centigrade. Active tuners may be required, or in the case of LEDA and the SNS RFQs, the vanes and the RFQ body are supplied with separate coolant temperatures [16] to differentially control the vane length from the body dimensions to hold the frequency constant.

## References

- [1] <http://laacg1.lanl.gov>
- [2] R. Duperrier et al, XX Int'l Linac Conf, p.839
- [3] B. Bondarev et al, PAC2001, p.2947
- [4] L. Young, PAC 1997, p.2752
- [5] R.A. Jameson, ORNL/TM-2007/001
- [6] I. Hofmann, PAC 1981, p.2399
- [7] D. Li et al, IPAC 2012
- [8] H. Vernon. Smith et al, PAC 2001, p.3297
- [9] [www.cst.de](http://www.cst.de)
- [10] [www.ansoft.com](http://www.ansoft.com)
- [11] K. Ko, Proc. ACES 2002 Conf, 2002
- [12] [www.ansys.com](http://www.ansys.com)
- [13] A. Pisent et al, Linac 2004, p.69
- [14] A. Pisent et al, EPAC 2008, p.3542
- [15] A. Pisent, Linac 2010, p.372
- [16] S. Virostek et al, Linac 2000, p.908

### 1.6.19 Rhodotron

*Y. Jongen, M. Abs, IBA*

The Rhodotron, invented by Jacques Pottier, CEA, 1987, is a novel accelerator principle suitable for electron acceleration up to 10-12 MeV. This new accelerator principle was the answer to the needs of high power electron beams of more than 2-3 MeV for industrial applications or for intense x-ray production. Energies below 3 MeV are still well served by electrostatic accelerators (Sec.1.6.5). At more than 3-5 MeV industrial linac's have been used so far but with a relatively poor electrical efficiency and limited beam control and intensity.

Ion Beam Applications, S.A. (Belgium) developed a range of 5 to 10 MeV Rhodotrons with average beam power ratings from 25 kW to 700 kW at working frequencies of 107 MHz and 215 MHz. Thirty machines have been sold so far.

**Beam acceleration** The Rhodotron is a recirculating rf accelerator based on a unique beam pattern that looks like a flower. See Figs.1 and 2. The beam crosses a number of times at different azimuths a half wave coaxial cavity in its median plane. Machines performing up to 12 crossings have been manufactured. This seems to be a practical limit. Each time the beam crosses the cavity it gains up to 1 MeV. The beam is then re-injected by external DC magnets towards the center of the cavity. The magnets' position and field are chosen in such that the beam is re-injected with the right phase with respect to the rf field. The first magnet is usually of a different size as the beam is not yet relativistic.

The big accelerating cavity has a fairly high Q factor (50,000 @ 107.5 MHz) and so has a relatively low power consumption thus enabling c.w. operation, even if pulse operation is possible. The central conductor of the cavity has two cones at the extremities thereby improving the Q and avoiding HOMs to be excited by the beam

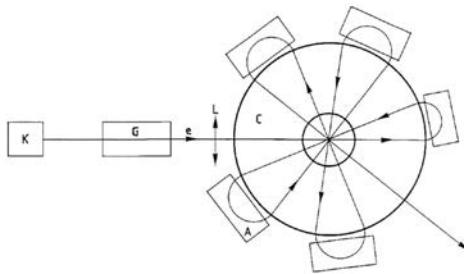


Figure 1: Horizontal section of the cavity showing the beam path and magnet positions.

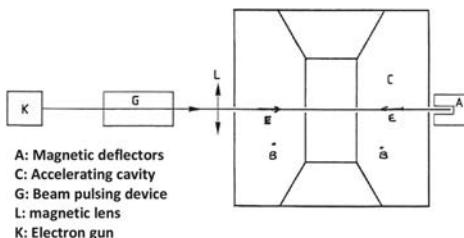


Figure 2: Vertical section showing the tapered central conductor.

harmonics. The cavity that is excited by an external power amplifier chain based on tetrodes does not require a tuning element. The small resonance frequency drifts due to thermal changes in the cavity dimensions are followed by the rf generator based on phase measurements on the final amplifier. An amplitude regulation is required to maintain the accelerating field stable at better than 1% as the beam power is changed.

**Beam injection** The injection of the beam into the cavity is done with an external electron gun. The injection energy is 30-60 keV. The electron train must be pulsed at working frequency to allow the beam transmission  $\sim 100\%$ . The typical phase acceptance is large and around  $60^\circ$ . This means that the injected peak current is 8 to 10 times the average current. Peak current up to 1 A, with average currents up to 100 mA, has been successfully injected in the most powerful model. The IBA e-guns are based on commercial cathode-grid assemblies that allow good beam control in time and amplitude. A beam current control accuracy of 0.1% is easily achieved.

**Beam focalization** Unlike linacs, the phase stability is of synchrotron type. This is due to the

fact that the magnets lengthen the beam trajectories and shift the phase when the energy increases.

The electromagnetic focusing forces due to the rf field have limited effects except on the first crossing. The vertical and horizontal focusing is ensured mainly by the magnets. No other elements are needed to control the beam size, making the machine simple and robust.

The extraction of the beam is obvious and allows placing different exits at different energies by selecting to switch off certain magnets.

### Machine properties summary

- Energy range: up to 10-12 MeV per accelerator (more cavities can be put in series to increase energy)
- Energy spread:  $< 300$  keV at 10 MeV
- Beam current control  $< 0.1\%$  precision in less than  $500 \mu\text{s}$
- Electrical efficiency: up to 55% at full beam power

### 1.6.20 Storage Rings

*W. Fischer, BNL*

Storage rings are circular machines that store particle beams at a constant energy. Beams are stored in rings without acceleration for a number of reasons (Tab.1). Storage rings are used in high-energy, nuclear, atomic, and molecular physics, as well as for experiments in chemistry, material and life sciences. Parameters for storage rings such as particle species, energy, beam intensity, beam size, and store time vary widely depending on the application. The beam must be injected into a storage ring but may not be extracted (Fig.1). Accelerator rings such as synchrotrons (Sec.1.6.21) are used as storage rings before and after acceleration.

Particles stored in rings include electrons and positrons; muons; protons and antiprotons; neutrons; light and heavy, positive and negative atomic ions of various charge states; molecular and cluster ions [1], and neutral polar molecules. Spin polarized beams of electrons, positrons, and protons were stored. The kinetic energy of the stored particles ranges from  $10^{-6}$  eV [2] to  $4.0 \times 10^{12}$  eV (LHC,  $7 \times 10^{12}$  eV planned), the number of stored particles from one (ESR [4]) to  $10^{15}$  (ISR [5]).

To store beam in rings requires bending (dipoles) and transverse focusing (quadrupoles).

Table 1: Storage ring applications with examples of past, existing, and planned machines.

|  |
|--|
| Beam accumulation: AA, AR, CR, EPA, MIMAS, PAR, PIA, PSR, RESR, RR, SNS                                    |
| Beam quality improvement: LEIR, ILC DR   |
| Stretcher, slow extraction: ELSA, KSR  |
| Synchrotron light source: ALS, APS, AS, BESSY, CLS, DIAMOND, ESRF, INDUS, NSLS, PLS, SOLEIL, SLS, Spring-8 |
| Collision with internal target: COSY, HERAE, HESR, IUCF, MIT-Bates, Nuclotron                              |
| Collider: AdA, BEPC, HERA, ISR, KEKB, RHIC, LHC, Tevatron, VEPP-2000                                       |
| Stored beam experiments: ASTRID, ESR, $g - 2$ , pEDM, TARN, TSR, UMER                                      |

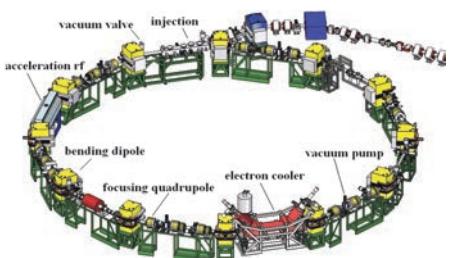


Figure 1: Small storage ring (CRYRING at the Manne Siegbahn Laboratory) with main components labeled.

Higher order multipoles are used to correct chromatic aberrations, to suppress instabilities, and to compensate for nonlinear field errors of dipoles and quadrupoles. Magnetic multipole functions can be combined in magnets. Beams are stored bunched with radio frequency (rf) systems, and unbunched. The magnetic lattice and rf system are designed to ensure the stability of transverse and longitudinal motion (Sec.1.6.21), respectively. New technologies allow for better storage rings. With strong focusing the beam pipe dimensions became much smaller than previously possible. For a given circumference superconducting magnets (Sec.7.2.4) make higher energies possible, and superconducting rf systems (Sec.7.3.9) allow for efficient replenishment of synchrotron radiation losses of large current electron or positron beams. Storage rings have instrumentation to monitor the electrical and mechanical systems, and the beam quality. Computers are used to control the operation (Sec.7.7). Large storage rings have millions of control points from all systems.

The time dependent beam intensity  $I(t)$  can often be approximated by an exponential function

$$I(t) = I(0) \exp(-t/\tau)$$

where the decay time  $\tau$  and, correspondingly, the store time ranges from a few turns to 13 days (ISR [6]).  $\tau$  can be dominated by a variety of effects including lattice nonlinearities (Sec.2.3), beam-beam (Sec.2.5), space charge (Sec.2.4.5.1), intrabeam and Touschek scattering (Sec.2.4.12), interaction with the residual gas or target (Sec.3.3.2.1), or the lifetime of the stored particle. In this case, the beam lifetime measurement itself can be the purpose of a storage ring experiment [1].

The main consideration in the design of a storage ring is the preservation of the beam quality over the store length. The beam size and momentum spread can be reduced through cooling (Sec.2.7), often leading to an increase in the store time. For long store times vacuum considerations are important since the interaction rate of the stored particles with the residual gas molecules is proportional to the pressure, and an ultra-high vacuum system may be needed (Sec.5.10). Distributed pumping with warm activated NEG surfaces or cold surfaces in machines with superconducting magnets are ways to provide large pumping speeds and achieve low pressures even under conditions with dynamic gas loads.

The largest application of storage rings today are synchrotron light sources (Sec.2.2.4), of which about 50 exist world wide. Storage ring light sources are continuously improved and will remain the dominant form for the foreseeable future [7].

In experiments where the beam collides with an internal target or another beam (Sec.1.6.3), a storage ring allows to reuse the accelerated beam many times if the interaction with the target is sufficiently small. In hadron colliders and ion storage rings store times of many hours or even days are realized, corresponding to up to  $10^{11}$  turns and thereby target passages. Ref.[3] is the first proposal for a collider storage ring.

A number of storage rings exist where the beam itself or its decay products are the object of study.

## References

- [1] M. Larsson, Rep. Prog. Phys. 58, p.1267 (1995)
- [2] K.-J. Kügler, W. Paul, U. Trinks, Phys. Lett. B, Vol.72, Issue 3, p.422 (1978)

## Sec.1.6: GLOSSARY OF ACCELERATOR TYPES

- [3] G. O'Neill, PR 102, 1418 (1956)
- [4] Yu.A. Litvinov et al, Nucl. Phys. A 756 3 (2005)
- [5] K. Johnsen, CERN 84-13 (1984)
- [6] W. Scharff, Particle Accelerators and their uses', Harwood (1986)
- [7] M. Bei et al, NIM A 622, 518 (2010)

### 1.6.21 Synchrotrons

*C. Zhang, S.X. Fang, IHEP*

Synchrotrons are characterized by the magnetic field synchronizing with momentum of particle beams and the electric field synchronizing with their circulating frequency, so that beams may circulate along stationary orbits and be continuously accelerated. Particle beams can be accelerated to higher momentum in synchrotrons, where the phase slipping is avoided for synchronization and magnets get less massive as in cyclotrons (Sec.1.6.4). In a synchrotron, the guide field  $B$  varies with time as

$$B(t) = \frac{p(t)c}{Zep} \quad (1)$$

where  $p$  is particle momentum,  $c$  is speed of light,  $\rho$  is orbit radius, and  $Ze$  denotes particle charge. As seen in Eq.(1),  $\rho$  stays constant if variation of  $B$  matches the increase in  $p$ .

Another requirement in synchrotrons is for the frequency of rf field  $f_{rf}$  to synchronize with circulating frequency  $f_0$  of the particle beam,

$$f_{rf} = h \cdot f_0 \quad (2)$$

Here  $h$  is harmonic number (integer). There may have up to  $h$  synchronizing particle bunches in a synchrotron.

Synchrotrons posed technical challenges to magnets, microwave as well as their control and these became possible after World War II. In 1946, F. Goward and D. Barnes at the Telecommunications Research Laboratory of UK modified its small betatron to operate as a synchrotron [1]. In the following year, a team at General Electric Co. constructed a dedicated 70 MeV electron synchrotron. In 1952, BNL finished a proton synchrotron of cosmic ray energy range, the 3 GeV Cosmotron.

Figure 1 shows the schematic of a synchrotron. As shown, particle beam is injected into the vacuum pipe through a deflector; bending and focusing magnets confine the particles to move along and around the central orbit; an rf cavity installed in a straight section accelerates the beam;

the magnetic field ramps along with the beam momentum increase; the ejection elements extract the beam when it reaches the desired energy.

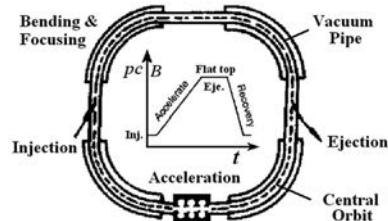


Figure 1: Schematic of a synchrotron.

As the energies and the arrival times to the rf cavity of the particles in a beam are slightly different, there are energy and phase deviations in reference to the idealized synchronous particle. The nonsynchronous particles will be lost in the accelerator if there is no longitudinal focusing. In 1944, Veksler [2] in the USSR and McMillan [3] in the US independently discovered the principle of phase stability in synchrotrons. Figure 2 illustrates the phase oscillation, or synchrotron oscillation, below and above the transition energy.

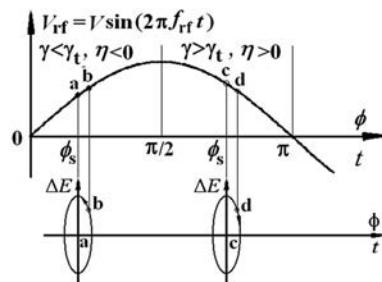


Figure 2: Phase oscillation in synchrotrons, stable phase is in the range  $(0, \pi/2)$  below the transition energy and  $(\pi/2, \pi)$  above the transition energy.

There are two factors affecting the revolution frequency of particles in a synchrotron, i.e. speed and orbit length: the higher the energy of a particle, the faster it moves and the longer is its orbit. The speed is dominating factor at low energy, while orbit length dominates when approaching the speed of light. This shows that there is a transition energy  $\gamma_t$ : higher energy particles circulate faster when  $\gamma < \gamma_t$ , and slower when  $\gamma > \gamma_t$ . Below transition, the particle "b" circulates with higher frequency for  $E_b > E_a$  and gains more

energy for  $\phi_b > \phi_a$  than the synchronized particle “a”, so that it oscillates around the particle “a” anticlockwise in the  $\phi$ - $\Delta E$  space when  $\gamma < \gamma_t$ . Similarly, particle “d” oscillates around the particle “c” clockwise when  $\gamma > \gamma_t$ . In case when  $\Delta\phi = \phi - \phi_s$  is small enough, the “restore force”,  $\Delta V_{rf} = V_{rf}(\phi) - V_{rf}(\phi_s) \approx 2\pi f_{rf} V \cos \phi_s \cdot \Delta\phi$ , is nearly constant and the synchrotron oscillation behaves as a simple pendulum,

$$\frac{d^2\Delta\phi}{dt^2} + \omega_s^2 \Delta\phi = 0 \quad (3)$$

with oscillation frequency

$$f_s = \frac{\omega_s}{2\pi} = \frac{f_0}{\beta_s} \sqrt{-\frac{h\eta \cos \phi_s}{2\pi} \frac{eV}{E_s}} \quad (4)$$

where  $\eta = 1/\gamma_t^2 - 1/\gamma^2$ ,  $\beta_s$  and  $E_s$  are relative velocity and energy of the synchronous particle respectively.

The transverse focusing was explored in cyclotrons (Sec.1.6.4). Particles get focusing in both horizontal and vertical planes if the guide field index  $n$  satisfies  $0 < n < 1$  with

$$n = -\frac{\rho}{B_0} \left( \frac{\partial B_y}{\partial r} \right)_{r=\rho} \quad (5)$$

However, this constant-gradient focusing is rather weak which makes the beam’s cross section large and leads to bulky vacuum chambers and massive magnets. The typical size of vacuum chamber is 800-mm wide by 200-mm height, while the total weight of magnates gets to  $\sim 10000$  tons in weak focusing synchrotrons of GeV scale such as the Cosmotron.

In the same year when Cosmotron was completed, E. Courant, M. Livingston and H. Snyder proposed the concept of strong focusing (or alternating gradient focusing) [4]. The idea had been visualized by Chritofilos in an unpublished paper [5]. The strong focusing opened a new era of synchrotrons, which allowed using smaller magnets and reaching higher beam energies. In a strong focusing synchrotron, horizontal focusing and defocusing (vertical focusing and defocusing) magnets are alternatingly arranged to make a global strong focusing like in an optical system,

$$\frac{1}{F} = \frac{1}{f_1} + \frac{1}{f_2} - \frac{d}{f_1 f_2} \quad (6)$$

If the lenses have equal and opposite focal lengths, i.e.  $f_1 = -f_2$ , the overall focal length  $F = d/f^2 > 0$ , with  $d$  distance between two lenses. There are two catalogues of strong focusing, i.e. combined function and separated function. In the former case, the bending and focusing

are combined in the same magnets, while in the latter case dipoles and quadrupoles play roles of bending and focusing separately, so that it is more flexible providing even stronger focusing.

The alternating gradient focusing quickly superseded constant focusing in synchrotron design. CERN immediately abandoned its 10 GeV weak focusing plan and constructed a 25 GeV strong focusing proton synchrotron. In 1959 and 1960, the CERN Proton Synchrotron (CPS), and the BNL’s Alternating Gradient Synchrotron (AGS) of 30 GeV completed successively. Their magnetic gaps are only 70-80 mm.

For the electric and magnetic fields to be modulated in synchronism with beam momentum, synchrotrons are operated in pulsed mode with typical average beam current of the order of  $\mu\text{A}$ ’s. The maximum beam current in synchrotrons is limited by space charge effects (Sec.2.4.5) and a variety of other instabilities, and also by beam-beam interactions in colliders (Sec.2.5).

Synchrotrons are applied for nuclear and particle physics experiments of fixed targets, and serve as booster injectors for higher energy accelerators. Synchrotrons are used to make collisions between oppositely directed beams, known as colliders, which in turn can provide high effective interaction energy (Sec.1.6.3). The high energy frontier accelerators constructed in recent years are all collider type. The light emitted from electron synchrotrons was harmful in the history of  $e^+e^-$  colliders, it has now become an important tool for scientific research (see also Sec.1.6.20). Dozens of synchrotron radiation sources have been constructed in the world. Proton synchrotrons are used as spallation neutron sources (Sec.1.6.16). Proton and heavy ion synchrotrons are also widely applied for medical treatment (Sec.1.6.14).

## References

- [1] F.K. Goward, D.E. Barnes, Nature, 158 (1946), 413
- [2] V.I. Veksler, Dokl. Akad. Nauk SSSR 43, 346 and 44, 393 (1944)
- [3] E.M. McMillan, PR 68 (1945) 143
- [4] E.D. Courant, M.S. Livingston, H.S. Snyder, PR 88 (1952)
- [5] N.C. Chrostofilos, U.S. Patent no. 2,736,799 (1956)

### 1.6.22 Two-Beam Accelerator

*J.P. Delahaye, CERN*

Novel schemes of Two-Beam Acceleration (TBA) have been proposed [1, 2] as power source for high-energy facilities and especially for  $e^\pm$  linear colliders. In the TBAs, a high-current, low-energy drive beam is used to generate rf power that is applied to a high-gradient acceleration structure, where a low-current beam is accelerated to high energy (Fig.1). Many variations of the TBA concept have been investigated, with early work [1] centered on using FELs to extract rf power from the drive beam. However, most rf extraction concepts (FELs, klystrons, gyrotrons, wake-field) can be configured into a TBA. There are also several choice options for the drive beam source, and the method of drive-beam acceleration. The TBA has the great advantage of high efficiency for power conversion from the drive beam to rf power. In addition, TBAs scale [3] favorably to high frequencies ( $\geq 11.4$  GHz) and high accelerating gradients ( $\geq 100$  MV/m).

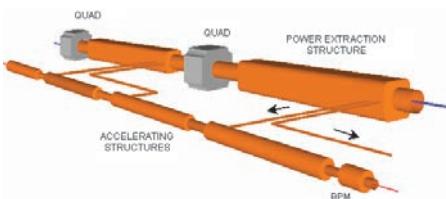


Figure 1: Conceptual layout of Two-Beam Accelerator (TBA).

Two main approaches of TBA research have been pursued so far:

**The relativistic klystron approach (RK-TBA)** [4] developed by LBNL/LLNL uses induction acceleration of the drive beam and to maintain the energy at 10 MeV throughout most of the device. A klystron-like output structure extracts power at 11.4 GHz from the drive beam. The primary technical challenge of the RK-TBA lies in propagating the intense drive beam (hundreds of amperes) at low energy (10 MeV) over long distances.

**The Compact Linear Collider (CLIC) scheme** [5, 6] is based on a relativistic drive beam which is not reaccelerated to avoid active elements in the main tunnel. The drive beam is characterized by a 12 GHz bunch structure, low energy (2.4 GeV) and high current (100 A). Successive drive beam trains supply power to a linac sector 876 m long.

Up to 90% of the beam energy is transformed in rf power after which the remaining beam is dumped. Acceleration of the main beam is pursued in a following sector using a fresh drive beam.

**Drive beam generation** The drive beam generation complex is shown in Fig.2 [7]. It produces one after the other all required drive beams for each linac. The drive beams are generated as one long train with a bunch spacing of 60 cm. A fully-loaded normal-conducting linac operating at a low frequency (1 GHz) is used to accelerate the drive beams to 2.4 GeV. The bunch spacing is then reduced to 2.5 cm in three successive stages in a delay loop and two combiner rings using funneling techniques to repetitively interleave 240 ns-long slices of the trains. As a result, the bunch repetition frequency and the beam intensity are multiplied by a factor of 24. Operating the linac in the fully-loaded mode enables the beam to be accelerated with an rf-power-to-beam efficiency of  $\approx 96\%$ . The rf power for each drive-beam accelerator is supplied by efficient 15 MW multi-beam klystrons with long rf pulse at low frequency. A particularly attractive and cost effective feature of the CLIC scheme is that energy upgrading of the collider only requires a pulse lengthening of the modulators which drive the klystrons and not an increase in the number of klystrons.

**Rf power production** By initially sending the drive beam trains in the opposite direction to the main beam, different time pulses in the train are used to power different sections of the main linac. The structures to extract the power from the high-intensity drive beam are referred to as PETS (Power Extraction and Transfer Structures) [8]. Each 0.5-m structure extracts a rf power of 130 MW from the 100 A drive beam. The structures are passive microwave devices in which the drive beam bunches interact with a large aperture (25 mm diameter) structure with a shallow periodically corrugated inner surface to preferentially excite the synchronous  $TM_{01}$  mode at 12 GHz. In the process, the beam kinetic energy is converted into electromagnetic energy at the mode frequency. This energy travels along the structure with the mode group velocity and the rf power produced is collected at the downstream end by a power extractor and conveyed to the main linac structure by rectangular waveguides. For the sake of beam stability, these structures have to be damped to reduce long-range transverse wakefield effects. PETS are made of eight octants separated

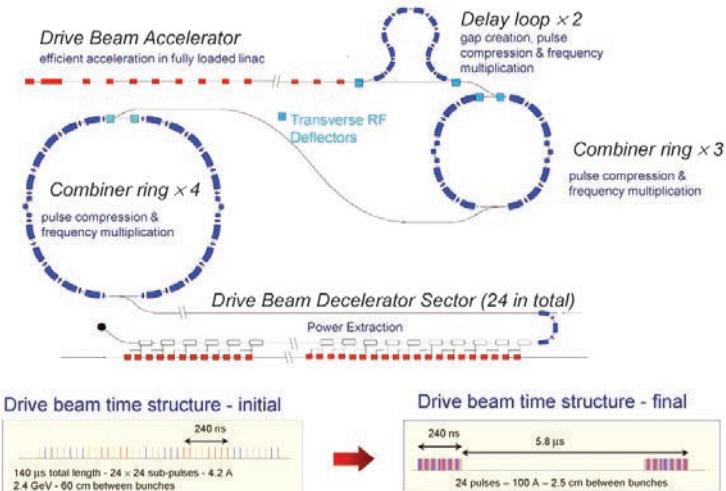


Figure 2: Layout of CLIC RF power generation scheme.

by HOM damping slots connected to broad-band SiC rf damping loads. In case of problems, the power generated by these structures is turned off by a remotely adjustable external reflector.

**Main beam acceleration** The main linac is made of a succession of two beam modules integrating all necessary components including rf structures, beam instrumentation, vacuum, quadrupoles, alignment & stabilization (Fig.3).

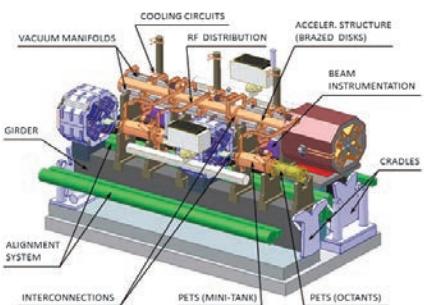


Figure 3: Two Beam Acceleration module.

**Experiments** The earliest TBA experiments [9] were performed on ETA-I addressing issues of power extraction. Work on the RK version started shortly after, using the ARC facility. These experiments [10] used a 1-MeV 1-kA 70-ns induction accelerator beam to produce 300 MW of rf power level at 11.4 GHz. Reacceleration experiments [11] were performed on ATA that demonstrated

bunched beam transport through two reacceleration induction cells and three traveling-wave extraction cavities. These experiments resulted in a total rf output of  $> 200$  MW, with phase and amplitude stable over a significant portion of the beam pulse. At LBNL a RK-TBA version [12] was designed as a power source for a linear collider with 1.5-TeV CM collision energy.

Two generations of CLIC Test Facilities (CTF1 1990-1995 and CTF2 1996-2002) have demonstrated the technical feasibility of the CLIC scheme. In CTF1, a peak power of 76 MW was extracted from the drive beam by a high-impedance 30-GHz traveling-wave section, and used to reaccelerate the same beam. In CTF2, a string of four power-extracting structures driving five accelerating structures increased the energy of a single electron bunch of the probe beam by 55 MeV [13]. Both the high intensity drive beam and the probe beam were generated by laser-illuminated photo-cathodes in rf guns. A new CLIC Test Facility CTF3 [14] has been built at CERN to address the major key CLIC-technology-related feasibility issues. In particular CTF3 has demonstrated the generation of a 130-ns 150-MeV 28-A drive beam with 2 cm bunch spacing using a fully-loaded linac and two stages of bunch interleaving resulting in an intensity and frequency multiplication by a factor 8 [15]. This beam will be used to produce up to 1.5 GW of 12 GHz power in a 20-m Test Beam Line decelerator in order to carry out beam stability studies

## Sec.1.7: ACCELERATOR COMPUTER CODES

and benchmark simulation codes. It will also be used to power accelerating structures at their nominal gradient of 100 MV/m in prototype Two Beam modules thus addressing the feasibility of Two Beam Acceleration [16].

### References

- [1] D.B. Hopkins, A.M. Sessler, J.S. Wurtele, NIM Phys. Res. 228 (1984) 15
- [2] A. Sessler et al, PRL 58, 2439 (1987); NIM A 306,592 (1991)
- [3] S.M. Lidia et al, Snowmass Workshop (1996)
- [4] T.L. Houck, G.A. Westenskow, IEEE Trans. On Plasma Sci. 24 (1996) 938
- [5] The CLIC Conceptual Design Report, <http://clic-study.org/accelerator/CLIC-ConceptDesignRep.php>
- [6] R. Tomas et al, PRST-AB 13,014801 (2010)
- [7] The CLIC RF Power Source, CLIC note 364
- [8] I. Syratchev, 7th Workshop on High Energy Density and High Power RF, Kalamata, Greece (2005)
- [9] D.B. Hopkins et al, SPIE High Intensity Laser Processes 664 (1986) 73
- [10] M.A. Allen et al, PRL 63 (1989) 2472
- [11] G.A. Westenskow, T.L. Houck, Proc.10th Int. Conf. High Energy Part. Beams (1994)
- [12] Zeroth-order Design Report for Next Linear Collider, Appendix A, SLAC-474 (1996) p.925
- [13] H.H. Braun, CERN-PS-2001-008
- [14] G. Geschonke et al, CTF3 design report, CERN/PS 2002-008
- [15] P. Skowronski et al, IPAC10
- [16] J.P. Delahaye et al, Mod. Phys. Lett. A, Vol.26, No.40, 2997 (2011)

### 1.7 ACCELERATOR COMPUTER CODES

*R. Ryne, LBNL*

Simulation is essential to accelerator design, commissioning, operation, and upgrades. At the present time several web sites maintain lists of codes used by various groups, including: [http://oraweb.cern.ch/pls/hhh/code\\_website disp\\_allcat/](http://oraweb.cern.ch/pls/hhh/code_website disp_allcat/), <http://www.slac.stanford.edu/accel/ilc/codes/>, [http://pbpl.physics.ucla.edu/Computing/Code\\_Overview/](http://pbpl.physics.ucla.edu/Computing/Code_Overview/), <http://www.cap.bnl.gov/mumu/>. In the future it is expected that a list of accelerator codes will be maintained at the website for Physical Review Special Topics Accelerators & Beams, <http://prst-ab.aps.org/>

### Community standards

- Accelerator Lattice Descriptions: Complex accelerator lattices may contain many thousands of elements, so community standards for lattice descriptions are important to facilitate sharing of lattice information among codes. An overview of accelerator description formats can be found at <http://cern.ch/AccelConf/ICAP06/PAPERS/THM2IS01.PDF>. The first, and most widely used, standard for the specification of lattices is the Standard Input Format (SIF); this is described in the documentation for the MAD-X code, <http://mad.web.cern.ch/mad>. The SIF definition of beamline elements has been very widely adopted. For realistic analysis (as contrasted with idealized design) two extensions of SIF have been found to be important: full instantiation (where every element has its own parameters, errors, etc.) and compatibility with modern software tools. The latter reduces mainly to standard eXtensible Markup Language (XML), preferably disciplined by an XML Schema; see, e.g., <http://code.google.com/p/ual/source/browse/trunk/doc/adxf>. A Universal Accelerator Parser has also been developed that parses several formats; see <http://www.lns.cornell.edu/~dcs/aml/>
- Self-Describing Data Sets (SDDS): SDDS is a self-describing file protocol that has been widely adopted in the light source modeling community; see [www.aps.anl.gov/Accelerator\\_Systems\\_Division/Operations\\_Analysis/SDDSInfo.shtml](http://www.aps.anl.gov/Accelerator_Systems_Division/Operations_Analysis/SDDSInfo.shtml)
- Experimental Physics and Industrial Control System (EPICS): EPICS is widely used in accelerator control systems. Given the ubiquity of EPICS, it is important to accelerator modelers who integrate on-line accelerator codes into control systems. Information about EPICS is located at <http://www.aps.anl.gov/epics/> and <http://www-csr.bessy.de/epics/>
- Radiation safety modeling and standards: See <http://rsicc.ornl.gov/index.html> and <http://irs.inms.nrc.ca/software/egsnrc/>

The following tables list some widely used (non-commercial) codes and contact information.

## Beam Dynamics Codes:

(Below, PIC refers to codes with particle-in-cell space-charge capability.)

| Code               | URL or Contact   | Description/Comments   |
|--------------------|--|--|
| ASTRA              | <a href="http://tesla.desy.de/~meykopff">tesla.desy.de/~meykopff</a>                             | 3D parallel, general charged particle beams incl. space charge   |
| AT                 | <a href="http://sourceforge.net/projects/atcollab/">sourceforge.net/projects/atcollab/</a>       | Accelerator Toolbox  |
| BETACOOL           | <a href="http://betacool.jinr.ru">betacool.jinr.ru</a>   | Long term beam dynamics: ECOOL, IBS, internal target             |
| Bmad, Tao          | <a href="http://www.lns.cornell.edu/~dcs/bmad/">www.lns.cornell.edu/~dcs/bmad/</a>               | General purpose toolbox library + driver program                 |
| COSY INFINITY      | <a href="http://www.cosyinfinity.org">www.cosyinfinity.org</a>                                   | Arbitrary-order beam optics code                                 |
| CSRTrack           | <a href="http://www.desy.de/xfel-beam/csrtrack">www.desy.de/xfel-beam/csrtrack</a>               | 3D parallel PIC; includes CSR; mainly for e- dynamics            |
| Elegant/SDDS suite | <a href="http://aps.anl.gov/elegant.html">aps.anl.gov/elegant.html</a>                           | parallel; track, optimize; errors; wakes; CSR                    |
| ESME               | <a href="http://www-ap.fnal.gov/ESME">www-ap.fnal.gov/ESME</a>                                   | Longitudinal tracking in rings                                   |
| HOMDYN             | <a href="mailto:Massimo.Ferrario@LNF.INFN.IT">Massimo.Ferrario@LNF.INFN.IT</a>                   | Envelope equations, analytic space charge and wake fields        |
| IMPACT code suite  | <a href="http://amac.lbl.gov">amac.lbl.gov</a>   | 3D parallel multi-charge PIC for linacs and rings                |
| LAACG code suite   | <a href="http://laacg.lanl.gov">laacg.lanl.gov</a>   | Includes PARMILA, PARMELA, PARMTEQ, TRACE2D/3D                   |
| LiTrack            | <a href="http://www.slac.stanford.edu/~emma/">www.slac.stanford.edu/~emma/</a>                   | Longitudinal linac dynamics; wakes; GUI-based; error studies     |
| LOCO               | <a href="mailto:safranek@slac.stanford.edu">safranek@slac.stanford.edu</a>                       | Analysis of optics of storage rings; runs under matlab           |
| LUCRETIA           | <a href="http://www.slac.stanford.edu/accel/ilc/codes">www.slac.stanford.edu/accel/ilc/codes</a> | Matlab-based toolbox for simulation of single-pass e- systems    |
| MaryLie            | <a href="http://www.physics.umd.edu/dsat">www.physics.umd.edu/dsat</a>                           | Lie algebraic code for maps, orbits, moments, fitting, analysis  |
| MaryLie/IMPACT     | <a href="http://amac.lbl.gov">amac.lbl.gov</a>   | 3D parallel PIC; MaryLie optics + IMPACT space charge            |
| MAD-X              | <a href="http://mad.web.cern.ch/mad">mad.web.cern.ch/mad</a>                                     | General purpose beam optics                                      |
| MERLIN             | <a href="http://www.desy.de/~merlin">www.desy.de/~merlin</a>                                     | C++ class library for charged particle accelerator simulation    |
| OPAL               | <a href="http://amas.web.psi.ch">amas.web.psi.ch</a>   | 3D parallel PIC; cyclotrons, FFAGs, linacs; particle-matter int. |
| ORBIT              | <a href="mailto:jzh@ornl.gov">jzh@ornl.gov</a>   | Collective beam dynamics in rings and transport lines            |
| PATH               | <a href="mailto:Alessandra.Lombardi@cern.ch">Alessandra.Lombardi@cern.ch</a>                     | 3D PIC; linacs and transfer lines; matching and error studies    |
| SAD                | <a href="http://acc-physics.kek.jp/SAD/sad.html">acc-physics.kek.jp/SAD/sad.html</a>             | Design, simulation, online modeling & control                    |
| SIMBAD             | <a href="http://agsrhighome.bnl.gov/People/luccio">agsrhighome.bnl.gov/People/luccio</a>         | 3D parallel PIC; mainly for hadron synchrotrons, storage rings   |
| SIXTRACK           | <a href="http://frs.home.cern.ch/frs/">frs.home.cern.ch/frs/</a>                                 | Single particle optics; long term tracking in LHC                |
| STRUCT             | <a href="http://www-ap.fnal.gov/users/drozhdin">www-ap.fnal.gov/users/drozhdin</a>               | Long term tracking w/ emphasis on collimators                    |
| Synergia           | <a href="https://compacc.fnal.gov/projects">https://compacc.fnal.gov/projects</a>                | 3d parallel PIC: space charge, nonlinear tracking and wakes      |
| TESLA              | <a href="mailto:lyyang@bnl.gov">lyyang@bnl.gov</a>   | Parallel; tracking; analysis; optimization                       |
| TRACK              | <a href="http://www.phy.anl.gov/atlas/TRACK">www.phy.anl.gov/atlas/TRACK</a>                     | 3D parallel PIC, mainly for ion or electron linacs               |
| LIBTRACY           | <a href="http://libtracy.sourceforge.net/">libtracy.sourceforge.net/</a>                         | Library for beam dynamics simulation                             |
| TREDI              | <a href="http://www.tredi.enea.it">www.tredi.enea.it</a>   | 3D parallel PIC; point-to-point Lienard-Wiechert                 |
| UAL                | <a href="http://code.google.com/p/ual/">code.google.com/p/ual/</a>                               | Unified Accelerator Libraries                                    |
| WARP               | <a href="http://DPGrote@lbl.gov">DPGrote@lbl.gov</a>   | 3D parallel ES and EM PIC with accelerator models                |
| ZGOUBI             | <a href="http://sourceforge.net/projects/zgoubi/">sourceforge.net/projects/zgoubi/</a>           | Magnetic optics; spin; sync radiation; in-flight decay           |

Beam Dynamics w/ emphasis on specific phenomena (beam-beam, e- -cloud, spin):

|              |  |  |
|--------------|--|--|
| BBSIM        | <a href="http://www-ap.fnal.gov/~tsen/BBSIM/index.html">www-ap.fnal.gov/~tsen/BBSIM/index.html</a>         | Beambeam simulations, compensation and diagnostics         |
| BBSS, BBWS   | <a href="mailto:ohmi@post.kek.jp">ohmi@post.kek.jp</a>   | Beam-beam strong-strong and weak-strong codes              |
| BBTRACK      | <a href="http://ab-abp-bbtrack.web.cern.ch/ab-abp-bbtrack/">ab-abp-bbtrack.web.cern.ch/ab-abp-bbtrack/</a> | Long-range beam-beam interaction studies                   |
| BeamBeam3D   | <a href="http://amac.lbl.gov">amac.lbl.gov</a>   | Parallel;strong-strong;multi-bunch;multi-IP;long-range     |
| BEAMX, COMBI | <a href="http://lhcb-beam-beam.web.cern.ch/lhc-beam-beam">lhcb-beam-beam.web.cern.ch/lhc-beam-beam</a>     | Parallel;strong-strong;FMM:symplectic 6D beam-beam         |
| C-MAD        | <a href="http://MPiVi@SLAC.Stanford.edu">MPiVi@SLAC.Stanford.edu</a>                                       | Parallel PIC self-consistent; e-cloud and instabilities    |
| ECLOUD       | <a href="http://ab-abp-rlc.web.cern.ch/ab-abp-rlc-ecloud">ab-abp-rlc.web.cern.ch/ab-abp-rlc-ecloud</a>     | e-cloud build-up; e- flux, heat load; multi-bunch wakes    |
| HEADTAIL     | <a href="http://ab-abp-rlc.web.cern.ch/ab-abp-rlc-ecloud">ab-abp-rlc.web.cern.ch/ab-abp-rlc-ecloud</a>     | Coll. effects modeling; head-tail; e- cloud instab; TMCI   |
| PEI, PEHTS   | <a href="mailto:ohmi@post.kek.jp">ohmi@post.kek.jp</a>   | cloud build-up, coupled bunch instab; head-tail instab     |
| POSINST      | <a href="mailto:MAFurman@lbl.gov">MAFurman@lbl.gov</a>   | 2D build-up code; detailed secondary emission model        |
| SPINK        | <a href="http://agsrhighome.bnl.gov/People/luccio">agsrhighome.bnl.gov/People/luccio</a>                   | Spin tracking of spin ½ particles                          |
| SPRINT       | <a href="mailto:Georg.Hoffstaetter@cornell.edu">Georg.Hoffstaetter@cornell.edu</a>                         | Spin tracking, incl. nonlin, invariant spin field, ramping |
| WARP/POSINST | <a href="mailto:JLVay@lbl.gov">JLVay@lbl.gov</a>   | Parallel 3D self-consistent e-cloud; mesh refinement       |

## Electromagnetics:

|                   |  |  |
|-------------------|--|--|
| ABC1              | <a href="http://abci.kek.jp">abci.kek.jp</a>   | 2.5D wake field computation code                 |
| ACE3P suite       | <a href="http://slacportals.slac.stanford.edu/sites/ard_public/bpd/acc">slacportals.slac.stanford.edu/sites/ard_public/bpd/acc</a> | 3D Parallel FE; Omega3P/S3P, Track3P,Pic3P,TEM3P |
| FEMAXX            | <a href="http://amas.web.psi.ch">amas.web.psi.ch</a>   | 3D Parallel FE Maxwell eigensolver               |
| MOEVE             | <a href="http://www.ief.uni-rostock.de/index.php?id=235">www.ief.uni-rostock.de/index.php?id=235</a>                               | Multi-grid Poisson solver                        |
| OSIRIS, QuickPIC  | <a href="http://exodus.physics.ucla.edu/codes.html">exodus.physics.ucla.edu/codes.html</a>   | EM and quasi-static particle-in-cell             |
| Poisson/Superfish | <a href="http://laacg.lanl.gov">laacg.lanl.gov</a>   | 2D magnet design and rf cavity design            |
| RADIA             | <a href="http://ftp.esrf.fr/pub/InsertionDevices/RADIA">ftp.esrf.fr/pub/InsertionDevices/RADIA</a>                                 | 3D magnetostatics; general magnet design         |
| ROXIE             | <a href="http://espace.cern.ch/roxie/default.aspx">espace.cern.ch/roxie/default.aspx</a>   | 3D magnetostatics; esp SC magnet design          |

## Sec.1.7: ACCELERATOR COMPUTER CODES

### Free Electron Laser Codes:

|         |                                     |   |
|---------|-------------------------------------|---|
| FAST    | mikhail.yurkov@desy.de              | 3D time-dependent FEL code  |
| GENESIS | svan.reiche@psi.ch                  | Parallel 3D time-dep FEL, harmonics                               |
| GINGER  | WMFawley@lbl.gov                    | r-z-t eikonal field solver; full 3D mover; amp. & osc. capability |
| MEDUSA  | Henry.P.Freund@saic.com             | Parallel 3D time-dep FEL amp & osc; non-wiggler averaged          |
| PERSEO  | www.perseo.enea.it                  | FEL-cad library; 1D simulation of SASE FEL                        |
| SIMPLEX | radiant.harima.riken.go.jp/simplex/ | 3-D FEL simulator, GUI-based                                      |

### Synchrotron Radiation Modeling Codes:

|                 |  |   |
|-----------------|--|---|
| Radiation2D     | www.shintakelab.com/en/enEducationalSoft.htm               | Radiation from accelerating charges         |
| SPECTRA         | radiant.harima.riken.go.jp/spectra                         | Spontaneous synchrotron radiation           |
| SPUR            | svan.reiche@psi.ch   | Lienard-Wiechert, spectral, incoherent rad. |
| SRW             | ftp.esrf.fr/pub/InsertionDevices/SRW                       | Synchrotron rad. from arbitrary source      |
| Synrad/Synrad3d | david.sagan@cornell.edu                                    | Synchrotron radiation in linacs and rings   |
| XOP             | www.esrf.eu/UsersAndScience/Experiments/TBS/SciSoft/xop2.3 | X-ray Oriented Programs                     |

### Beam/Material Interactions:

|            |                                |  |
|------------|--------------------------------|--|
| FLUKA      | www.fluka.org                  | Particle transport and interactions with matter    |
| GEANT4     | www.geant4.org/geant4/         | Toolkit for passage of particles through matter    |
| G4Beamline | g4beamline.muonsinc.com        | Particle tracking in beamlines using GEANT4        |
| ICCOOL     | pubweb.bnl.gov/~fernnow/icool/ | Ionization cooling channel design                  |
| MARS       | www-ap.fnal.gov/MARS/          | 3D particle transport and interactions with matter |
| MCNP       | mcnp-green.lanl.gov/index.html | General purpose Monte Carlo transport code         |

### Pre-, Post-Processors, Parsers, Auxiliary Codes:

|         |                               |   |
|---------|-------------------------------|---|
| AML/UAP | www.lns.cornell.edu/~dcs/aml/ | XML-based Accel. Markup Language, Universal Accel. Parser   |
| H5hut   | h5hut.psi.ch/, vis.lbl.gov/   | HDF5 Utility Toolkit, parallel I/O for particles and fields |
| Sussix  | emaclean.web.cern.ch/emaclean | Resonance driving terms and frequency map analysis          |

**Other Web Resources** Since 2006, the proceedings of the International Computational Accelerator Physics (ICAP) conference series have been administered by the Joint Accelerator Conferences Website (JACoW); they are accessible and searchable at <http://www.jacow.org>. As of this writing, Physical Review Special Topics - Accelerators and Beams is planning to be a resource for information about accelerator codes; see <http://prst-ab.aps.org/>. Another searchable resource is the CERN Document Server, <http://cdsweb.cern.ch>.

## Errata and Additions to Handbook, 2nd Edition

### January 18, 2016

p.170, left column, Dynamic emittance

Original text

New text A more complete expression for the dynamic emittance was computed in:

- A.V.Otboyev, E.A.Perevedentsev, PRST-AB 2, 104401 (1999)

p.261, second row from the top, longitudinal impedance and wake of a BPM consisting of two strip lines

Original text  $Z_0^{\parallel} = 2Z_c \left[ \frac{\phi_0}{2\pi} \right]^2 [2 \sin^2 kL - i \sin 2kL]$ ,  $W'_0 = 2Z_c c \left[ \frac{\phi_0}{2\pi} \right]^2 [\delta(z) - \delta(z+2L)]$

New text  $Z_0^{\parallel} = Z_c \left[ \frac{\phi_0}{2\pi} \right]^2 [2 \sin^2 kL - i \sin 2kL]$ ,  $W'_0 = Z_c c \left[ \frac{\phi_0}{2\pi} \right]^2 [\delta(z) - \delta(z+2L)]$

p.572, right column, Lifetimes of carbon foils

Original text  $k_{\text{foil}} = 0.018$  for foils produced by vapor deposition

New text  $k_{\text{foil}} = 0.0018$  for foils produced by vapor deposition

p.576, right column, References

Original text

New text Remove Ref.[1]. Add two additional references:

- L.R. Scherk, Can. J. Phys. 57, 558 (1979)
- P.B. Keating et al., Phys. Rev. A 52, 4547 (1995)

# Chapter 2. BEAM DYNAMICS

## 2.1 PHASE SPACE

### 2.1.1 Linear Betatron Motion [1, 2]

D.A. Edwards, DESY

M.J. Syphers, Michigan State U.

#### Equations of motion

$$\begin{aligned} x'' + K_x x = 0, \quad K_x &\equiv \frac{e}{p} \frac{\partial B}{\partial x} + \frac{1}{\rho^2} \\ y'' + K_y y = 0, \quad K_y &\equiv -\frac{e}{p} \frac{\partial B}{\partial x} \\ z' &= -x/\rho \end{aligned} \quad (1)$$

where  $x, y$  are the transverse displacements from the reference particle of another particle of equal momentum  $p$ , and  $z$  the distance by which it leads the reference particle, which is itself a distance  $s$  along the reference trajectory from the origin. The reference trajectory is flat in the  $x, s$  plane, with its bending radius of curvature  $\rho$  due only to a static magnetic field in the  $y$  direction.

**Solution in phase-amplitude form** Below when results are expressed for  $x, x'$ , typically they apply for  $y, y'$  as well.

$$x = A\sqrt{\beta} \cos(\psi + \delta) \quad (2)$$

$$x' = -\frac{A}{\sqrt{\beta}} [\alpha \cos(\psi + \delta) + \sin(\psi + \delta)]$$

where  $A$  and  $\delta$  are constants of integration,  $\alpha \equiv -\frac{1}{2}d\beta(s)/ds$ .

The *amplitude function*, or  $\beta$ -function, satisfies

$$2\beta\beta'' - \beta'^2 + 4\beta^2K = 4 \quad (3)$$

and the phase advances according to  $d\psi/dx = 1/\beta$ . Expressing  $A$  in terms of  $x, x'$  yields

$$A^2 = \gamma x^2 + 2\alpha xx' + \beta x'^2 \quad (4)$$

$$= \frac{1}{\beta} [x^2 + (\alpha x + \beta x')^2]$$

with  $\gamma \equiv (1 + \alpha^2)/\beta$ . In a single pass system such as a linac, the *Courant-Snyder parameters*

$\alpha, \beta, \gamma$  may be selected to match the  $x x'$  distribution of the input beam. In a recursive system, the parameters are usually defined by the structure rather than by the beam.

#### Matrix parameterization

$$\begin{aligned} \begin{bmatrix} x_2 \\ x'_2 \end{bmatrix} &= \mathbf{M}(s_1, s_2) \begin{bmatrix} x_1 \\ x'_1 \end{bmatrix} \quad (5) \\ m_{11} &= \sqrt{\frac{\beta_2}{\beta_1}} (\cos \Delta\psi + \alpha_1 \sin \Delta\psi) \\ m_{12} &= \sqrt{\beta_1 \beta_2} \sin \Delta\psi \\ m_{21} &= -\frac{1 + \alpha_1 \alpha_2}{\sqrt{\beta_1 \beta_2}} \sin \Delta\psi + \frac{\alpha_1 - \alpha_2}{\sqrt{\beta_1 \beta_2}} \cos \Delta\psi \\ m_{22} &= \sqrt{\frac{\beta_1}{\beta_2}} (\cos \Delta\psi - \alpha_2 \sin \Delta\psi) \end{aligned}$$

When  $s_1$  and  $s_2$  are separated by the repetition length of the system, e.g. circumference  $C$  in a synchrotron,  $\mathbf{M}$  reduces to

$$\mathbf{M}_C = \mathbf{I} \cos \psi_C + \mathbf{J} \sin \psi_C, \quad \mathbf{J} = \begin{bmatrix} \alpha & \beta \\ -\gamma & -\alpha \end{bmatrix} \quad (6)$$

Note that  $\mathbf{J}^2 = -\mathbf{I}$  and so  $\mathbf{M}_C^n = \mathbf{I} \cos n\psi_C + \mathbf{J} \sin n\psi_C$ . The transformation of  $\mathbf{J}$  from one location to another is given by  $\mathbf{J}_2 = \mathbf{M}(s_1, s_2) \mathbf{J}_1 \mathbf{M}(s_1 s_2)^{-1}$ , the associated phase advance is

$$\Delta\psi = \tan^{-1} \left( \frac{m_{12}}{\beta_1 m_{11} - \alpha_1 m_{12}} \right) \quad (7)$$

where the  $m_{ij}$  are elements of  $\mathbf{M}$ . The condition for stable motion is  $|\cos \mu| = \frac{1}{2} \operatorname{tr} \mathbf{M}_C \leq 1$ .

**Emittance** The characterization of the area occupied by the beam in  $x, x'$  space or  $y, y'$  space is termed the *emittance* in that degree-of-freedom. Various definitions are in use. For a Gaussian beam with standard deviation  $\sigma$  in  $x$  and  $\beta x' + \alpha x$ , the area containing a fraction  $F$  of the beam is

$$\varepsilon = -\frac{2\pi\sigma^2}{\beta} \ln(1 - F) \quad (8)$$

## Sec.2.1: PHASE SPACE

A more general definition which reduces to the 15% case of Eq.(8) is

$$\varepsilon = \sqrt{\langle x^2 \rangle \langle x'^2 \rangle - \langle xx' \rangle^2} \quad (9)$$

and the beam “sigma” matrix is defined by

$$\Sigma = \begin{bmatrix} \langle x^2 \rangle & \langle xx' \rangle \\ \langle xx' \rangle & \langle x'^2 \rangle \end{bmatrix} = \varepsilon \begin{bmatrix} \beta & -\alpha \\ -\alpha & \gamma \end{bmatrix} \quad (10)$$

The *normalized emittance*,  $\varepsilon_N \equiv \gamma(v/c)\varepsilon$ , where here  $\gamma$  is the Lorentz factor, is proportional to the area in  $x, p_x$  phase space and so is expected to exhibit adiabatic invariance under energy change.

**Momentum dispersion** Trajectories of particles having momenta different from that of the reference trajectory by  $\Delta p$  contain the additional term  $D(s)\Delta p/p_s$ , where  $D$  is the *dispersion function*.

$D$  satisfies an inhomogeneous Hill’s equation

$$D'' + \left( K_x \frac{p_s}{p} - \frac{1}{\rho^2} \frac{\Delta p}{p} \right) D = \frac{1}{\rho} \frac{p_s}{p} \quad (11)$$

where  $\rho$  is the radius of curvature for  $p_s$ .

The matrix  $M(s_1, s_2)$  for the bend plane may be enlarged to include propagation of off-momentum rays,

$$\begin{bmatrix} x_2 \\ x'_2 \\ z_2 \\ \frac{\Delta p}{p_s} \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} & 0 & m_{14} \\ m_{21} & m_{22} & 0 & m_{24} \\ m_{31} & m_{32} & 1 & m_{34} \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x'_1 \\ z_1 \\ \frac{\Delta p}{p_s} \end{bmatrix} \quad (12)$$

where the elements of Eq.(5) are unchanged and

$$m_{14} = D_2 - m_{11}D_1 - m_{12}D'_1 \quad (13)$$

$$m_{24} = D'_2 - m_{21}D_1 - m_{22}D'_1$$

$$m_{31} = D'_1 + m_{21}D_2 - m_{11}D'_2$$

$$m_{32} = -D_1 + m_{22}D_2 - m_{12}D'_2$$

$$m_{34} = -D_2(m_{21}D_1 + m_{22}D'_1)$$

$$+ D'_2(m_{11}D_1 + m_{12}D'_1) + \frac{s_2 - s_1}{\gamma^2}$$

Particles differing in momenta will follow paths of differing lengths. For a sufficiently long path  $S$ , e. g. the circumference of a large synchrotron, with the use of Eqs.(12-13) the relative path length difference  $\Delta S$  becomes

$$\frac{\Delta S}{S} = -\frac{z_2 - z_1}{s_2 - s_1} = \left[ \left\langle \frac{D}{\rho} \right\rangle - \frac{1}{\gamma^2} \right] \frac{\Delta p}{p_s} \quad (14)$$

The *compaction factor*,  $\alpha_p$ , and the *transition gamma*,  $\gamma_t$ , are defined by

$$\alpha_p = \frac{1}{\gamma_t^2} \equiv \left\langle \frac{D}{\rho} \right\rangle \quad (15)$$

**Tune and chromaticity** The number of oscillations per turn of circumference  $C$  in a cyclic accelerator is the *tune*,

$$\nu \equiv \frac{\Delta\psi_C}{2\pi} = \frac{1}{2\pi} \oint \frac{ds}{\beta} \quad (16)$$

The momentum dependence of restoring force leads to a tune variation characterized by the *chromaticity*,  $\xi$ ,

$$\Delta\nu = \xi \frac{\Delta p}{p_s} \quad (17)$$

The *natural chromaticity* associated with the linear optics is

$$\xi = -\frac{1}{4\pi} \oint K\beta ds \quad (18)$$

A single sextupole magnet of length  $\ell$  and  $B'' \equiv \partial^2 B_y / \partial x^2$  will contribute

$$\Delta\xi = \pm \frac{1}{4\pi} D\beta \frac{B''\ell}{(B\rho)} \quad (19)$$

where the + sign is associated with the bend plane; this is the basis for chromaticity adjustment using sextupoles.

## References

- [1] E.D. Courant, H.S. Snyder, Ann. Phys. 1 (1958) 1
- [2] D.A. Edwards, M.J. Syphers, An Introduction to the Physics of High Energy Accelerators, Wiley (1993)

### 2.1.2 Longitudinal Motion [1, 2]

D.A. Edwards, DESY  
M.J. Syphers, Michigan State U.

**Equations of motion** The only role of the transverse degrees-of-freedom taken into account in this section is the bend-plane momentum dispersion function  $D$  defined in Sec.2.1.1, which leads to changes in  $z$ . Momentum change is due to an RF electric field  $\mathcal{E}$  in the longitudinal direction acting on charges of magnitude  $e$ ,

$$z' \equiv \frac{dz}{ds} = -\frac{D(s)}{\rho(s)p_s} \Delta p \quad (1)$$

$$\Delta p' \equiv \frac{d\Delta p}{ds} = \frac{e\mathcal{E}(s+z) - e\mathcal{E}(s)}{v_s} \quad (2)$$

**Bunch compression** The principle of a bunch compressor may be illustrated by a magnetic chicane of four magnets with bends  $\theta, -\theta, -\theta, \theta$  separated by distances  $L, S, L$ . The result is  $\Delta z = -2L\theta\Delta p/p_s$  in the limit of magnets of zero length. If the particle momentum increases from bunch front to rear, the bunch becomes shorter in passing through this system.

**Synchrotron oscillations** Here it is assumed that the rate of energy change is sufficiently low so that the electric field may be represented by a traveling wave traveling with the reference particle. The difference equation approach suitable to abrupt energy change is discussed in [2]. In the slow case, Eqs.(1) and (2) become

$$z' = -\frac{\eta}{p_s} \Delta p, \quad \eta = \frac{1}{\gamma_t^2} - \frac{1}{\gamma^2} \quad (3)$$

$$\Delta p' = \frac{e\mathcal{E}_0}{v_s} [\sin(ks + kz) - \sin(ks)] \quad (4)$$

where  $k \equiv 2\pi/\lambda$ ; for a cyclic accelerator the wavelength  $\lambda$  is the circumference divided by an integer  $h$ . Transition gamma,  $\gamma_t$ , is defined in Sec.2.1.1. Combining Eqs.(3) and (4) gives

$$z'' + \frac{p'_s}{p_s} z' + \frac{\eta e \mathcal{E}_0}{p_s v_s} [\sin(ks + kz) - \sin(ks)] = 0 \quad (5)$$

For sufficiently small  $z$  and  $p'_s = 0$ ,

$$z'' + \frac{2\pi\eta e V_0}{p_s v_s h \lambda^2} \cos(ks) z = 0 \quad (6)$$

where  $V_0 = \mathcal{E}_0 h \lambda$ . The number of oscillations in a distance  $h\lambda$  – the synchrotron oscillation tune – is

$$v_s = \left( \frac{h\eta}{2\pi\beta^2 E_s} e V_0 \cos ks \right)^{1/2} \quad (7)$$

where  $E_s$  is the energy of the reference particle and here  $\beta \equiv v_s/c$ . So, for stable oscillations above transition energy it is necessary that  $\cos ks > 0$ . This condition differs in sign from that in [1, 2] due to the use here of  $z$  rather than the phase at rf station passage. A first integral of Eq.(5) for  $p'_s = 0$  is

$$\frac{1}{2} \Delta p^2 + \frac{1}{2} U = C \quad (8)$$

where  $C$  is a constant, and

$$U = \frac{-1}{\pi h \eta} \frac{e V_0 p_s}{v_s} [\cos k(s+z) + kz \sin ks] \quad (9)$$

The “potential energy”  $U$  has zero slope at  $z = z_1 \equiv (\lambda/2 - 2s_{\text{mod}}\lambda)$ ; this is the unstable fixed point in the neighborhood of which the differential equation approach becomes invalid. The stable region in longitudinal phase space is termed a *bucket*; note that  $U$  has the form of a descending series of buckets above transition and an ascending series below.

**Bucket area and longitudinal emittance** The separatrix passes through the unstable fixed points located at  $z_1$  and the turning point  $z_2$  found from  $\cos k(s+z_2) + kz_2 \sin ks = \cos k(s+z_1) + kz_1 \sin ks$ . The area bounded by the separatrix may be calculated by numerical integration. In the particular case of  $\sin ks = 0$ , a *stationary bucket*,  $z_2 = -z_1 = -\lambda/2$ , and the bucket area is

$$\mathcal{A}_0 = \frac{8\lambda}{\pi c} \sqrt{\frac{E_s e V_0}{2\pi h \eta}} \quad (10)$$

The stable range of  $z$  is that for which  $\Delta p/p_s$  does not exceed the bucket height

$$\left( \frac{\Delta p}{p_s} \right)_{\max} = \sqrt{\frac{2eV_0}{\pi h \eta p_s v_s}} \quad (11)$$

A *longitudinal emittance* may be defined using Eq.(8) in analogy with its transverse counterpart, and in  $z$ ,  $\Delta p$  coordinates is expected to exhibit adiabatic invariance with energy.

The intrinsic nonlinearity of Eq.(2) can be ameliorated through the use of harmonic RF structures. An example may be found in the installation at DESY of a superconducting 3.9 GHz structure in the 1.3 GHz-based FLASH free electron laser [3].

## References

- [1] E.D. Courant, H.S. Snyder, Ann. Phys. 1 (1958) 1
- [2] D.A. Edwards, M.J. Syphers, An Introduction to the Physics of High Energy Accelerators, Wiley (1993)
- [3] E.R. Harms et al, Proc. SRF 2009, <http://accel-conf.web.cern.ch/AccelConf/srf2009/index.htm>

### 2.1.3 Linear Coupled System

D.A. Edwards, DESY

M.J. Syphers, Michigan State U.

Linear coupling between two of the three degrees-of-freedom may for the purposes of this section be characterized by a  $6 \times 6$  matrix of the form

$$\mathbf{M} = \begin{bmatrix} A & D & F \\ E & B & 0 \\ G & 0 & C \end{bmatrix} \quad (1)$$

where  $A-G$  are  $2 \times 2$  matrices.  $\mathbf{M}$  operates on a column vector  $(x, x', y, y', z, \Delta p/p)$  or  $(x, p_x, y, p_y, z, \Delta p)$ . With either choice  $\mathbf{M}$  is symplectic (Sec.2.3.6); the second, related to the canonical formalism, may be preferable if conservation of phase space volume is of interest.

**Transverse coupling** The equations of motion:

$$\begin{aligned} x'' + K_x(s)x &= S(s)y + R(s)y' + \frac{1}{2}R(s)'y \\ y'' + K_y(s)y &= S(s)x - R(s)x' - \frac{1}{2}R(s)'x \\ S &\equiv B'_{skew}/(\beta\rho), \quad R \equiv B_s/(\beta\rho) \\ B' &\equiv \frac{\partial B_y}{\partial x}, \quad B'_{skew} \equiv \frac{\partial B_x}{\partial x}, \quad R' \equiv \frac{\partial R}{\partial s} \end{aligned} \quad (2)$$

The  $S$  and  $R$  terms arise from skew quadrupole and solenoidal fields. Skew quadrupole fields are common in synchrotrons due to magnet fabrication or alignment errors.

A perturbative treatment of the transverse oscillations  $x = a(\beta_x/\beta_0)^{1/2} \cos \psi_x$ ,  $y = b(\beta_y/\beta_0)^{1/2} \cos \psi_y$  shows that at the sum resonance  $\nu_x + \nu_y = \text{integer}$ ,  $a^2 - b^2 = \text{constant}$ , so the sum resonance leads to instability [1]. In contrast, for the difference resonance  $a^2 + b^2 = \text{constant}$  and the motion though coupled is stable.

A potential benefit of complete coupling for an electron synchrotron is discussed in Sec.2.2.5. A positive use of solenoidal coupling for phase space manipulation is found in the round-to-flat beam transformation [2, 3, 4].

**Transverse-longitudinal coupling** In a synchrotron, the frequency modulation of the betatron oscillations arising from the natural chromaticity,  $\xi$ , leads to sidebands  $\nu_x \pm n\nu_z$ , an instance of synchrobetatron resonances (Sec.2.3.4). The natural chromaticity is usually compensated by the introduction of sextupole magnets; this step in itself implies the irreversible introduction of nonlinearities into the dynamics. Also discussed in Sec.2.3.4 are the sidebands due to passage through accelerating structures in which the dispersion  $D \neq 0$ , as well as the sidebands associated with the transverse fields,  $\mathcal{E}_x$  and  $\mathcal{B}_y$ .

That these same fields can be used to advantage is illustrated by the transverse-longitudinal emittance interchange process, in which a deflecting mode rf structure is placed at the midpoint of a 4 bend double “dogleg” arrangement [5]. For a TM<sub>110</sub>  $\pi$ -mode pillbox structure the equations of motion for this process are

$$\begin{aligned} x' &= \frac{p_x}{p_s}, \quad p'_x = \frac{e\mathcal{E}'}{c} \cos(ks) \cdot z \\ z' &= \frac{e\mathcal{E}'}{p_s c k} \sin(ks) \cdot x \\ p'_z &= \frac{e\mathcal{E}'}{c} \cos(ks) \cdot x + \frac{e\mathcal{E}'}{kc} \sin(ks) \cdot x' \end{aligned} \quad (3)$$

where the fields are those of the vector potential  $\mathcal{A}_s = -(\mathcal{E}'/\omega) \sin(\omega t) \cdot x$ . The reference particle is at the midpoint of each cell at the zero-crossing of the field, and  $\mathcal{E}'$  is the amplitude of  $\partial\mathcal{E}/\partial x$  at  $x = 0$ . For an  $n$ -cell structure the matrix is

$$\begin{bmatrix} 0 & 0 & -\frac{1}{\alpha} - \frac{n\lambda}{4D} & -\frac{n\alpha\lambda}{4} \\ 0 & 0 & -\frac{1}{D} & -\alpha \\ -\alpha & -\frac{n\alpha\lambda}{4} & 0 & 0 \\ -\frac{1}{D} - \frac{1}{\alpha} - \frac{n\lambda}{4D} & 0 & 0 & 0 \end{bmatrix} \quad (4)$$

where  $\alpha$  is the magnitude of the bend angle of each of the four magnets that create the dispersion  $D$  after the first two, and each cell of the pillbox cavity has length  $\lambda/2$ . The matrix used for the cavity is limited to lowest order in  $\lambda/(2\pi nD)$ , and the exchange is ideal as in [5]. Experimental results cited in [6] were made with a 5-cell structure of TESLA style. This process is also finding use in beam diagnostics [7]. See also Sec.2.2.5.

## References

- [1] E.D. Courant, H.S. Snyder, Ann. Phys. 1 (1958) 1
- [2] R.A. Brinkmann, Ya. Derbenev, K. Flöttmann, PRST-AB 4, 053501 (2001)
- [3] Yin-e Sun, PhD thesis, U. Chicago (2005), Fermilab-thesis-2005-17, <http://lss.fnal.gov/archive/thesis/fermilab-thesis-2005-17.shtml>
- [4] A.W. Chao, Proc. 11th Chinese Accel. Phys. symp., Fuzhou (2011)
- [5] K.J. Kim, A. Sessler, AIP Conf. Proc. Vol.821, p.115 (2005)
- [6] T.W. Koeth, PhD thesis, Rutgers U. (2009), <http://hdl.rutgers.edu/1782.2/rucore10001600001.ETD.000051364>
- [7] Y-E. Sun et al, <http://arxiv.org/abs/1003.3126>

### 2.1.4 Orbital Eigen-Analysis for Electron Storage Ring

J.A. Ellison, U. New Mexico

H. Mais, DESY

G. Ripken, Deceased 2004

**a. Basic problem** A general 6-D formalism is presented for the calculation of the bunch parameters (e.g. 6-D stationary beam-envelope matrix) for electron storage rings including radiation damping and quantum excitation. The problem is formulated in terms of a stochastic differential equation (SDE) and basic to our approach is the orbital eigen-analysis first introduced in [1]. The latter gives a more general framework than that of

Courant-Synder. An SDE approach was first introduced in [2] and developed further in [3]. At a later stage work [4] was incorporated and the starting point is the SDE in Frenet-Serret coordinates with respect to a design orbit as described in [5]. The 6D-SDE is

$$\vec{x}' = \mathbf{A}(s)\vec{x} + \vec{c}(s) + \epsilon\vec{g}(\vec{x}, s) + \epsilon^{\frac{1}{2}}\sqrt{b(s)}\xi(s)\vec{e}_6 \quad (1)$$

where  $\vec{x} = (x, p_x, y, p_y, z, p_z)^t$ . We have expanded up to second order in the dynamical variables and retained only the leading nonlinearities due to sextupoles and due to radiation effects in quadrupoles. All functions except  $\xi$  are  $C$ -periodic in  $s$  where  $C$  is the ring circumference. The matrix  $\mathbf{A}(s)$  is Hamiltonian and gives rise to the linear symplectic synchrobetatron oscillations. The  $\vec{x}$ -independent term  $\vec{c}$  describes closed orbit distortions induced by dipole field errors and by the fact that the energy losses in the bending magnets and quadrupoles are not replaced at the location in the ring where they occur. Our analysis is perturbative. The parameter  $\epsilon$  is inserted to indicate the perturbation size and to discuss the perturbation procedure and the nature of the error estimates. In the end,  $\epsilon$  can be taken to be one in applications of the formulas. The first perturbation term is  $\vec{g}(\vec{x}, s) = \delta\mathbf{A}(s)\vec{x} + \vec{f}(\vec{x}, s)$ . Here  $\delta\mathbf{A}$  models both the energy losses from radiation and the energy gain in the rf cavities, and  $\vec{f}$  takes into account the nonlinear terms due to sextupoles and due to radiation effects in quadrupoles. The last term in (1) simulates the stochastic excitation of the particle motion due to the quantum nature of the radiation. Here  $\xi$  is Gaussian white noise,  $b$  is an amplitude function proportional to  $\hbar$  and  $\vec{e}_k$  is the unit vector with 1 in the  $k$ -component, thus the stochastic excitation only affects the  $p_z$  component directly. The explicit form of these quantities can be found in [5] and details of our analysis below will be given in [6].

The main quantity of interest is the  $N$  particle random bunch density

$$\rho_N(\vec{x}, s) := \frac{1}{N} \sum_{n=1}^N \delta(\vec{x} - \vec{x}_n(s)) \quad (2)$$

where the  $\vec{x}_n(s)$  are independent and identically distributed random variables determined by (1). Let  $p$  be the single particle probability density defined by (1), then  $\langle \rho_N \rangle = p$ . Here, and in the following, angular brackets will denote the expected value of stochastic quantities. We will assume for large  $N$  that  $\rho_N(\vec{x}, s) \approx p(\vec{x}, s)$ , in a coarse

grained sense. This article presents results of an analysis of this single particle probability density. More precisely, the (periodic) 6-D closed orbit  $\vec{x}_{co}$ , satisfying (1) with  $b = 0$  will be defined by an integral equation, (1) will be linearized around  $\vec{x}_{co}$  and the linearized equation analyzed.

### **b. The equation $\vec{x}' = \mathbf{A}(s)\vec{x}$ and its Eigen-FSM**

The solutions of the linear periodic Hamiltonian system

$$\vec{x}' = \mathbf{A}(s)\vec{x}, \mathbf{A}^t \mathbf{J} + \mathbf{J} \mathbf{A} = 0, \mathbf{A}(s+C) = \mathbf{A}(s) \quad (3)$$

are central to our analysis. Here  $\mathbf{J} = \text{diag}(\mathbf{J}_2, \mathbf{J}_2, \mathbf{J}_2)$  is the unit symplectic matrix where  $\mathbf{J}_2 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ .

The transfer map,  $\mathbf{M}(s, s_0)$ , is a fundamental solution matrix (FSM) in  $s$  which satisfies  $\mathbf{M}(s_0, s_0) = I$  (often called the principal solution matrix). The basic properties are (P1)  $\mathbf{M}(s_2, s_1)\mathbf{M}(s_1, s_0) = \mathbf{M}(s_2, s_0)$  (transport property), (P2)  $\mathbf{M}^t \mathbf{J} \mathbf{M} = \mathbf{J}$  (symplecticity) and (P3)  $\mathbf{M}(s+C, s_0+C) = \mathbf{M}(s, s_0)$  (periodicity).

We assume all solutions of (3) are bounded. This is the case if and only if the monodromy matrix,  $\mathbf{M}(C, 0)$ , has six linearly independent eigenvectors,  $\vec{w}_k$ , and its eigenvalues (*characteristic multipliers*),  $\rho_k = \exp(i2\pi\nu_k)$ , have modulus one where the  $\nu_k$  are tunes (*characteristic exponents*), see [7]. To avoid resonance, we further assume that the  $\rho_k$  are distinct. Because  $M$  is real we can choose the  $\nu_k$  such that

$$0 < \nu_1 < \nu_3 < \nu_5 < 1/2, \quad \nu_{2l} = -\nu_{2l-1}, \quad (4)$$

and the  $\vec{w}_k$  such that  $\vec{w}_{2l} = \vec{w}_{2l-1}^*$ . Thus  $(1 - \rho_j^* \rho_k)\vec{w}_j^H \mathbf{J} \vec{w}_k = 0$ , so that  $\vec{w}_j^H \mathbf{J} \vec{w}_k = 0$  for  $j \neq k$ . Here  $H$  denotes Hermitian conjugate. Let  $\vec{w}_k = \vec{a}_k + i\vec{b}_k$ ,  $\vec{a}_k$  and  $\vec{b}_k$  real then  $\vec{w}_k^H \mathbf{J} \vec{w}_k = i\gamma_k$  where  $\gamma_k = 2\vec{a}_k^H \mathbf{J} \vec{b}_k$  and  $\gamma_{2l} = -\gamma_{2l-1}$ . More compactly

$$\mathbf{W}^H \mathbf{J} \mathbf{W} = i\tilde{\mathbf{I}}, \quad \text{where} \quad (5)$$

$$\mathbf{W} = [\vec{w}_1, \dots, \vec{w}_6], \quad \tilde{\mathbf{I}} = \text{diag}(\gamma_1, \dots, \gamma_6).$$

Finally, the  $\gamma_k$  are nonzero (since  $\mathbf{W}$  is nonsingular) and we normalize the  $\vec{w}_{2l-1}$  so that  $\gamma_{2l-1} = \pm 1$  (it appears the sign is not known a priori).

The phase space density of a bunch can be efficiently approximated in terms of the Eigen-FSM for (3),

$$\Psi(s) := \mathbf{M}(s, 0)\mathbf{W} \quad (6)$$

which was first introduced in [1] and generalizes the Courant-Synder formalism. It follows from (5) and (P2) that

$$\Psi^H(s)\mathbf{J}\Psi(s) = i\tilde{\mathbf{I}} \quad (7)$$

Define  $\hat{\Psi}$  by

$$\Psi(s) =: \hat{\Psi}(s) \exp(i\omega s N), \quad \omega := 2\pi/C \quad (8)$$

where  $N = \text{diag}(\nu_1, \dots, \nu_6)$ , then  $\hat{\Psi}$  is  $C$ -periodic and (8) is a Floquet representation of the Eigen-FSM. Furthermore  $\hat{\Psi}$  satisfies (7). The  $k$ -th column of  $\hat{\Psi}$  will be denoted by  $\vec{\psi}_k$ .

**c. Closed orbit and associated linearized SDE**

Since no characteristic multiplier,  $\rho_k$ , of (3) is one, (1), with  $b = 0$ , has a unique  $C$ -periodic solution for  $\epsilon$  sufficiently small. This closed orbit solution is defined by the integral equation

$$\begin{aligned} \vec{x}_{co}(s, \epsilon) &= \vec{x}_{co}(s, 0) \\ &+ \epsilon \int_0^C \mathbf{G}(s, t) \vec{g}(\vec{x}_{co}(s+t, \epsilon), s+t) dt \end{aligned} \quad (9)$$

Here  $\mathbf{G}(s, t) = (\mathbf{M}(s, s+C) - \mathbf{I})^{-1} \mathbf{M}(s, s+t)$  and  $\vec{x}_{co}(s, 0) = \int_0^C \mathbf{G}(s, t) \vec{c}(s+t) dt$ . See Thm. 2.1, p.154 of [8]. The closed orbit can be determined approximately by iterating (9).

Let  $\vec{y} := \vec{x} - \vec{x}_{co}$  then for  $\vec{y}$  small,  $\vec{y} \approx \vec{y}_L$  where  $\vec{y}_L$  satisfies the linear SDE

$$\vec{y}'_L = (\mathbf{A}(s) + \epsilon \mathbf{B}(s)) \vec{y}_L + \epsilon^{\frac{1}{2}} \sqrt{b(s)} \xi(s) \vec{e}_6 \quad (10)$$

Here  $\mathbf{B}(s) := D_1 \vec{g}(\vec{x}_{co}(s), s)$  is the Jacobian matrix of  $\vec{g}(\cdot, s)$ . The most important information about the bunch is contained in the moment (beam-envelope) matrix  $\mathbf{U}(s) = \langle \vec{y}_L(s) \vec{y}_L^T(s) \rangle$ . We now determine an approximation to  $\mathbf{U}$ .

**d. Equation for moment matrix and averaging approximation** The transformation  $\vec{y}_L \rightarrow \vec{z}$  via  $\vec{y}_L =: \Psi(s) \vec{z}$  gives

$$\begin{aligned} \vec{z}' &= \epsilon \mathbf{D}(s) \vec{z} + \epsilon^{\frac{1}{2}} \xi(s) \vec{d}(s) \\ \vec{d}(s) &:= \sqrt{b(s)} \Psi^{-1}(s) \vec{e}_6 \\ \mathbf{D}(s) &:= \Psi^{-1}(s) \mathbf{B}(s) \Psi(s) \end{aligned} \quad (11)$$

Now  $\mathbf{U} = \Psi \mathbf{V} \Psi^H$  where  $\mathbf{V} = \langle \vec{z} \vec{z}^H \rangle$  and the differential equation for  $\mathbf{V}$  is

$$\mathbf{V}' = \epsilon [\mathbf{D}(s) \mathbf{V} + \mathbf{V} \mathbf{D}^H(s) + \mathbf{E}(s)] \quad (12)$$

where  $\mathbf{E}(s) = \vec{d}(s) \vec{d}^H(s)$ .

Applying averaging methodology [9, 6] to (12) we obtain  $\mathbf{V}(s) \approx \mathbf{V}_a(s)$  where

$$\mathbf{V}'_a = \epsilon [\bar{\mathbf{D}} \mathbf{V}_a + \mathbf{V}_a \bar{\mathbf{D}}^H + \bar{\mathbf{E}}] \quad (13)$$

Here  $\bar{\mathbf{D}}$  and  $\bar{\mathbf{E}}$  denote the average of the quasiperiodic functions  $\mathbf{D}(s)$  and  $\mathbf{E}(s)$ , e.g.,

$$\bar{\mathbf{D}} = \lim_{L \rightarrow \infty} \frac{1}{L} \int_0^L \mathbf{D}(s) ds$$

Because of the non-resonant tune condition (4), the averages  $\bar{\mathbf{D}}$  and  $\bar{\mathbf{E}}$  are diagonal with

$$\bar{\mathbf{D}}_{jj} = \lambda_j := -i \gamma_j \vec{\psi}_j^H(s) \mathbf{J} \mathbf{B}(s) \vec{\psi}_j(s) \quad (14)$$

$$\bar{\mathbf{E}}_{jj} = \overline{b(s) |\hat{\Psi}(s)_{5j}|^2} \quad (15)$$

where the overbar denotes the  $s$  average of the underlying periodic function.

The solution of the IVP for (13) is

$$\begin{aligned} \mathbf{V}_a(s)_{jk} &= \exp[\epsilon(\lambda_j + \lambda_k^*)s] (\mathbf{V}_0 - \mathbf{V}_{ae})_{jk} \\ &\quad + \mathbf{V}_{ae\,jk} \\ \mathbf{V}_{ae\,jk} &= -\delta_{jk} \bar{\mathbf{E}}_{jj} / 2\Re \lambda_j \end{aligned} \quad (16)$$

Here  $\mathbf{V}_0$  is the initial condition for (12) and  $\mathbf{V}_{ae}$  is the unique equilibrium solution of (13).

In summary, the stochastic process  $\vec{x}$ , defined by (1), is given in a linear approximation by  $\vec{x}_{co} + \vec{y}_L$  and the moment matrix of  $\vec{y}_L$  is  $\mathbf{U} = \Psi \mathbf{V} \Psi^H$ , where the averaging approximation to the moment matrix is

$$\mathbf{U}(s) \approx \mathbf{U}_a(s) := \Psi(s) \mathbf{V}_a(s) \Psi^H(s) \quad (17)$$

**e. Dissipative case** For  $\Re \lambda_j < 0$ , the dissipative case,  $\mathbf{V}_a(s) \rightarrow \mathbf{V}_{ae}$  as  $s \rightarrow \infty$  and the approximate moment matrix in (17) becomes  $\mathbf{U}_a(s) = \Psi(s) \mathbf{V}_{ae} \Psi^H(s) = \hat{\Psi}(s) \mathbf{V}_{ae} \hat{\Psi}^H(s)$ . Using (14), (15) and (16), this stationary,  $C$ -periodic  $\mathbf{U}_a$  can be written

$$\begin{aligned} \mathbf{U}_a(s)_{jk} &= \sum_{l=1,2,3} G_l \Re[(\vec{\psi}_{2l-1}(s))_j (\vec{\psi}_{2l-1}^*(s))_k] \\ G_l &= -\frac{2}{\alpha_l} \frac{1}{C} \int_0^C b(s) |\vec{\psi}_{2l-1}(s)_5|^2 ds \end{aligned} \quad (18)$$

The quantities  $\alpha_l := 2\Re \lambda_{2l-1}$  are called the damping constants and are given by

$$\begin{aligned} \Re \lambda_{2l-1} &= \Re \lambda_{2l} \\ &= \frac{\gamma_{2l-1}}{C} \int_0^C \Im[\vec{\psi}_j^H(s) \mathbf{J} \mathbf{B}(s) \vec{\psi}_j(s)] ds < 0 \end{aligned}$$

**f. Robinson sum rule** Since  $\mathbf{D}$  and  $\mathbf{B}$  are related by the similarity transformation in (11),  $\text{Tr} \mathbf{D}(s) = \text{Tr} \mathbf{B}(s)$  and thus the  $C$ -periodicity of  $\mathbf{B}$  gives

$$\text{Tr} \bar{\mathbf{D}} = \frac{1}{C} \int_0^C \text{Tr} D_1 g(x_{co}(s), s) ds = \frac{2U_0}{E_0} \quad (19)$$

where  $U_0$  and  $E_0$  are the energy gain in the cavity and beam energy respectively (See [5]). From (14),  $\lambda_{2l}^* = \lambda_{2l-1}$  so that  $\sum_1^3 \alpha_l = \text{Tr} \bar{\mathbf{D}}$  and thus (19) is the Robinson sum rule [10].

## g. Remarks

- The averages of  $\mathbf{E}$  and  $\mathbf{D}$  were computed under the nonresonance condition of (4). However the standard averaging error bound  $|\mathbf{V}(s) - \mathbf{V}_a(s)| < O(\epsilon)$  for  $0 \leq s < O(1/\epsilon)$  requires a sufficient,  $O(1)$ , separation between  $0, \nu_1, \nu_3, \nu_5$ , and  $1/2$ . In the dissipative case, the  $s$ -interval of validity of the averaging approximation can be extended to all  $s \geq 0$ . Details are given in [6]. The resonant case is considered in [11].
- The mean of  $\vec{y}_L$  is easily handled and in the averaging approximation  $\langle \vec{y}_L(s) \rangle \approx \Psi(s) \exp(\epsilon \mathbf{D}s) \Psi^{-1}(0) \langle \vec{y}_L(0) \rangle$  and the covariance matrix is easily computed. Of course, in the dissipative case the mean approaches zero for large  $s$ . If  $\vec{y}_L(0)$  is a Gaussian random vector, then  $\vec{y}_L$  is a Gaussian process.
- Spin-orbit motion in electron storage rings and especially spin diffusion due to quantum fluctuations can be treated approximately by introducing an 8-D matrix formalism (Secs.2.6.7, 2.6.8). The codes SLIM and SLICK discussed in Sec.2.6.7 are based on the orbital eigen-analysis of this section and can be used to calculate beam polarization, as well as the orbital dynamics of this section.
- A general formalism for treating the linear electron beam dynamics with radiation effects taken into account is also presented in [12] and [13]. Whereas [12] starts from a kinetic description (Fokker-Planck equation), [13] uses (as we do) the SDE for the particle motion. The second order moments (beam envelopes) are calculated directly (rather than using the orbital-FSM) and these results are used in the computer code SAD. Furthermore generalized radiation integrals are derived which in the limiting case of a completely uncoupled machine reduce to the well known results of Sands, [14] (Sec.3.1.4.1). This is also true in our case if we separate the six dimensional dynamics into fast betatron and slow synchrotron components via the dispersion (Sec.2.6.8). See [15] for a discussion of the beam envelope without radiation effects.

## References

- [1] A.W. Chao, J. Appl. Phys. 50, 595 (1979) and EPAC 2008
- [2] A.A. Kolomensky, A.N. Lebedev, Theory of cyclic accelerators, North-Holland, 1966
- [3] H. Mais, G. Ripken, DESY 83-062 (1983)
- [4] J. Jowett, SLAC-Pub-4033 (1986)
- [5] D.P. Barber et al, DESY M-94-09 (1994)
- [6] J.A. Ellison, H. Mais, K. Heinemann, Details of Orbital Eigen-analysis for Electron Storage Rings, in preparation (to be posted on arXiv).
- [7] K.R. Meyer et al, Introduction to Hamiltonian Dynamical Systems and the N-Body Problem, Springer, NY, 2009
- [8] J.K. Hale, Ordinary Differential Equations, Krieger, Florida, 1980
- [9] J.A. Sanders et al, Averaging Methods in Nonlinear Dynamical Systems, 2nd Ed., Springer, 2007; J.A. Ellison, H.-J. Shih, AIP Conf. Proc. 326, 1995
- [10] K.W. Robinson, PR 111, 373 (1958)
- [11] J. Wu et al, PAC 2005; B. Nash et al, PRST-AB 9, 032801 (2006)
- [12] F. Ruggiero et al, Ann. Phys., 197, 396 (1990)
- [13] K. Ohmi et al, PRE, 49, 751 (1994)
- [14] M. Sands, SLAC-121 (1970)
- [15] A. Wolski, PRST-AB 9, 024001 (2006)

## 2.2 OPTICS AND LATTICES

### 2.2.1 Single Element Optics

*K. Brown, Deceased 2002*

**Charged particle optics** The following of a charged particle through a system of magnetic lenses may be represented by matrix multiplication. At any position  $s$  measured along a reference trajectory, a charged particle is represented by a vector (single column matrix)  $\mathbf{X}(s)$  with  $\mathbf{X}^t(s) = (x(s), x'(s), y(s), y'(s), l(s), \Delta P/P_0)$ , where  $x$  = the horizontal displacement of a trajectory (with respect to the central trajectory). The  $x$  plane is typically defined as the midplane of a bending magnet;  $x' = dx/ds$  is the angle this trajectory makes in the horizontal plane;  $y$  = the vertical displacement;  $y' = dy/ds$  is the vertical angle;  $l$  = the path length difference between the trajectory and the central trajectory as measured from the beginning of the system,  $s = 0$ ;  $\Delta P/P_0$  is the momentum deviation of the trajectory from that of the central trajectory. The central orbit is defined as that trajectory whose initial conditions

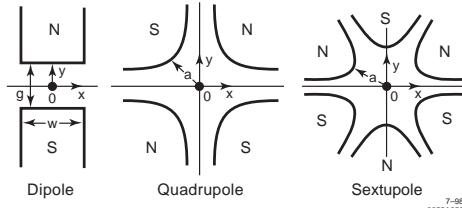


Figure 1: Magnet profiles with midplane symmetry.

are  $\mathbf{X}(0) = 0$ . The momentum of the central trajectory is  $P_0$ . Fig.1 shows a few magnet cross-sections with midplane symmetry.

The vector  $\mathbf{X}(0)$  at  $s = 0$  is transformed to another vector  $\mathbf{X}(1)$  at position 1 by

$$\mathbf{X}(1) = \mathbf{R}\mathbf{X}(0) \quad (1)$$

where  $\mathbf{R}$  is a  $6 \times 6$  matrix characterizing each magnet or drift distance.

All physics is contained in  $\mathbf{R}$  representing each type of magnet and  $\mathbf{R}$  is derived from the basic equations of motion of a particle passing through a static magnetic field. The traversing of several magnets and interspersing drift spaces is described by the same matrix equation (1) with a total  $\mathbf{R}$  given by

$$\begin{aligned} \mathbf{R}(s) &= \mathbf{R}(n) \cdots \mathbf{R}(i) \cdots \mathbf{R}(2)\mathbf{R}(1) \\ \implies \mathbf{X}(s) &= \mathbf{R}(s)\mathbf{X}(0) \end{aligned} \quad (2)$$

This formalism may be extended to second- and higher-order terms [1, 2].

Eq.(1) can be expanded as  $\mathbf{X}(s) = \mathbf{R}\mathbf{X}(0) =$

$$\begin{bmatrix} R_{11} & R_{12} & R_{13} & R_{14} & 0 & R_{16} \\ R_{21} & R_{22} & R_{23} & R_{24} & 0 & R_{26} \\ R_{31} & R_{32} & R_{33} & R_{34} & 0 & R_{36} \\ R_{41} & R_{42} & R_{43} & R_{44} & 0 & R_{46} \\ R_{51} & R_{52} & R_{53} & R_{54} & 1 & R_{56} \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x(0) \\ x'(0) \\ y(0) \\ y'(0) \\ l(0) \\ \Delta P/P_0 \end{bmatrix}$$

For static magnetic fields,  $P_0$  and  $\Delta P/P_0$  are constants of the motion. The determinant  $\det \mathbf{R} = 1$ . The 0 elements in row six occur because  $\Delta P/P_0$  is a constant of the motion in static magnetic fields. The 0 elements in column five are because  $x, x', y, y'$  and  $\Delta P/P_0$  do not depend upon  $l$ .

If all magnets in the system have midplane symmetry about  $y = 0$ , then in first order optics the  $x$ - and  $y$ -motions are decoupled, and

$$\mathbf{R} = \begin{bmatrix} R_{11} & R_{12} & 0 & 0 & 0 & R_{16} \\ R_{21} & R_{22} & 0 & 0 & 0 & R_{26} \\ 0 & 0 & R_{33} & R_{34} & 0 & 0 \\ 0 & 0 & R_{43} & R_{44} & 0 & 0 \\ R_{51} & R_{52} & 0 & 0 & 1 & R_{56} \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad (3)$$

Note that some of the  $\mathbf{R}$  matrix elements have a specific physical meaning,

$$R_{21} = -1/F_x, \quad R_{43} = -1/F_y \quad (4)$$

where  $F_{x,y}$  are the focal lengths of the system in the  $x$  and  $y$  planes; and  $R_{16}$  and  $R_{26}$  are the momentum dispersion and the angular dispersion in the  $x$  plane, respectively.

**Ray tracing using the R matrix** The  $\mathbf{R}$  matrix allows one to trace individual trajectories (rays) through any systems of magnets. To track a beam of particles (assume Gaussian distribution) through the system, the beam may be characterized by a “beam ellipse”. The equation of an ellipse in  $n$ -D may be formulated by the matrix equation

$$\mathbf{X}^t \Sigma^{-1} \mathbf{X} = 1 \quad (5)$$

where  $\Sigma$  is a positive definite  $n \times n$  moment matrix. For example, in 2-D,

$$\mathbf{X} = \begin{bmatrix} x \\ x' \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{21} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}$$

$$\implies \sigma_{22}x^2 - 2\sigma_{21}xx' + \sigma_{11}x'^2 = \det \Sigma \quad (6)$$

See Fig.2.

The area of the ellipse is

$$\text{Area} = \pi(\det \Sigma)^{1/2}$$

$$= \pi x_{\max} x'_{\max} = \pi x_{\max} x'_{\max} = \pi \epsilon \quad (7)$$

where  $\sqrt{\sigma_{11}} = x_{\max}$ ,  $\sqrt{\sigma_{22}} = x'_{\max}$ ,  $\epsilon$  = the emittance of the beam, and  $r_{21} = \sigma_{21}/\sqrt{\sigma_{11}\sigma_{22}}$  is the correlation between  $x$  and  $x'$ .

The transformation of the beam ellipse from position  $s = 1$  to a position  $s = 2$  in a beam line is given by

$$\Sigma_2 = \mathbf{R}\Sigma_1\mathbf{R}^t \quad (8)$$

The same equation applies for the  $y, y'$  plane and also for the 6-D ellipsoid where  $\sqrt{\sigma_{11}} = x_{\max}$ ,  $\sqrt{\sigma_{22}} = x'_{\max}$ ,  $\sqrt{\sigma_{33}} = y_{\max}$ ,  $\sqrt{\sigma_{44}} = y'_{\max}$ ,  $\sqrt{\sigma_{55}} = l_{\max}$ ,  $\sqrt{\sigma_{66}} = (dP/P_0)_{\max}$ , and  $r_{ij} = \sigma_{ij}/\sqrt{\sigma_{ii}\sigma_{jj}}$  is the correlation between the  $i$  and  $j$  variables in the 6-D phase space. As a result of the fact that the  $\Sigma$  matrix is positive definite, the  $r_{ij}$  fall in the range  $-1$  to  $+1$ .

**Courant-Snyder notation** (See also Sec.2.1.1)

The  $\Sigma$  matrix notation is rewritten as

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{21} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} = \epsilon \mathbf{T}, \quad \mathbf{T} \equiv \begin{bmatrix} \beta & -\alpha \\ -\alpha & \gamma \end{bmatrix} \quad (9)$$

with  $\det \mathbf{T} = 1$ . The equation of the ellipse (Courant-Snyder invariant) is then generated by the matrix equation

$$\mathbf{X}^t \mathbf{T}^{-1} \mathbf{X} = \epsilon, \quad \text{or} \quad \gamma x^2 + 2\alpha xx' + \beta x'^2 = \epsilon \quad (10)$$

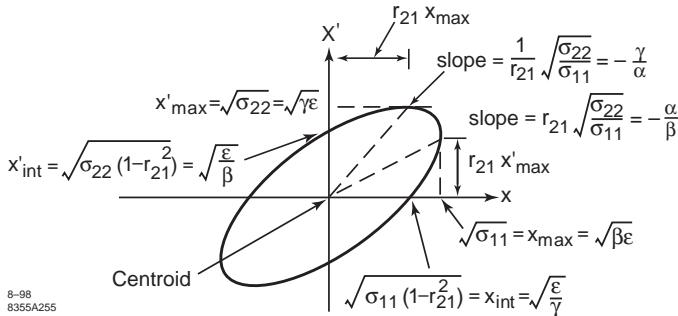


Figure 2: A two-dimensional beam phase ellipse.

The area of the ellipse is  $A = \pi\epsilon$  and the transformation of the ellipse parameters from position 1 to position 2 is

$$\mathbf{T}_2 = \mathbf{R}\mathbf{T}_1\mathbf{R}^t \quad (11)$$

which gives the transformation of  $(\beta, \alpha, \gamma)$ .

### R matrices for important optical elements

Drift space  $L$  = The length of the drift space.

$$\mathbf{R}_{\text{drift}} = \begin{bmatrix} 1 & L & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & L & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad (12)$$

Thick lens quadrupole  $\mathbf{R}_{\text{thick quad}} =$

$$\begin{bmatrix} \cos kL & \frac{\sin kL}{k} & 0 & 0 & 0 & 0 \\ -ksinkL \cos kL & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \cosh kL & \frac{\sinh kL}{k} & 0 & 0 \\ 0 & 0 & ksinkhL \cosh kL & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad (13)$$

This represents a quadrupole which is focusing in the  $x$  plane and defocusing in the  $y$  plane. This  $\mathbf{R}$  matrix is measured from the input face of the quadrupole to the exit face.  $L$  = the effective magnetic length of the quadrupole;  $a$  = the radius of the aperture of the quadrupole;  $B_0$  = the magnetic field strength at the radius  $a$ ;  $k^2 = (B_0/a)(1/B\rho)$ , where  $(B\rho)$  is the magnetic rigidity (momentum) of the central reference trajectory,  $(B\rho) = 33.3564 P_0$  kG-m when  $P_0$  is in GeV/c.

Thin lens quadrupole The two principal planes for a thin lens quad are both located at the center of

the thick quad. Then, with  $L$  = thick quad length,  $\mathbf{R}_{\text{thin quad, pp}} =$

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ -k \sin kL & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & k \sinh kL & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad (14)$$

Note that for both the thick lens and thin lens,  $R_{21} = -1/F_x$  and  $R_{43} = -1/F_y$ .

Wedge bending magnet  $\mathbf{R}_{\text{wedge bend}} =$

$$\begin{bmatrix} \cos k_x L & \frac{1}{k_x} \sin k_x L & 0 & 0 & 0 & 0 \\ -k_x \sin k_x L & \cos k_x L & 0 & 0 & 0 & 0 \\ 0 & 0 & \cos k_y L & 0 & 0 & 0 \\ 0 & 0 & -k_y \sin k_y L & 0 & 0 & 0 \\ \frac{h}{k_x} \sin k_x L & \frac{h}{k_x^2} (1 - \cos k_x L) & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{h}{k_x^2} (1 - \cos k_x L) & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{h}{k_x} \sin k_x L & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{h}{k_x} \sin k_x L \\ \frac{1}{k_y} \sin k_y L & 0 & 0 & 0 & 0 & 0 \\ \cos k_y L & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & \frac{h^2}{k_x^3} (k_x L - \sin k_x L) & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad (15)$$

where  $\rho_0$  = the bending radius of the central trajectory;  $h = 1/\rho_0$ ,  $k_x^2 = (1-n)h^2$ ,  $k_y^2 = nh^2$ ,  $\alpha = hL$  = the angle of bend of the central orbit,  $L$  = the path length of the central orbit,  $n = -(dB/dx)(\rho_0/B_0)$  is the normalized field gradient of the bending magnet measured on the optical axis ( $x = 0, y = 0$ ),  $R_{21} = -1/F_x$ ,  $R_{43} = -1/F_y$ .

Most of the time uniform field ( $n = 0$ ,  $k_x = h = 1/\rho_0$ ,  $k_y = 0$ ) magnets are used

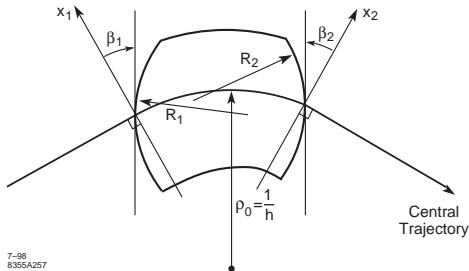


Figure 3: A bending magnet. The sign conventions for  $x, \beta, R$  and  $h$  are all positive as shown in the figure. Positive  $y$  is out of the paper. Positive  $\beta$ 's imply transverse focusing. Positive  $R$ 's (convex curvatures) represent negative sextupole components of strength  $S = (-h/2R) \sec^3 \beta$  [1]. A wedge bending magnet corresponds to  $\beta_1 = \beta_2 = 0$ .

for optical systems. Then for wedge magnets,  $\mathbf{R}_{\text{uniform wedge bend}} =$

$$\begin{bmatrix} C & S/h & 0 & 0 & 0 & (1-C)/h \\ -hS & C & 0 & 0 & 0 & S \\ 0 & 0 & 1 & L & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ S & (1-C)/h & 0 & 0 & 1 & (\alpha - S)/h \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad (16)$$

where  $C = \cos \alpha$ ,  $S = \sin \alpha$ . If the  $\mathbf{R}$  matrix is measured at the center of the wedge magnet, i.e. between the principal planes of the wedge magnet (Fig.4), then it becomes simple: Here we separate the  $\mathbf{R}_x$  and  $\mathbf{R}_y$  matrices and leave out the path length terms,

$$\begin{aligned} \mathbf{R}_{x,\text{pp}} &= \begin{bmatrix} 1 & 0 & 0 \\ -hS & 1 & S \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ x' \\ \Delta P/P_0 \end{bmatrix} \\ \mathbf{R}_{y,\text{pp}} &= \begin{bmatrix} 1 & L \\ 0 & 1 \end{bmatrix} \begin{bmatrix} y \\ y' \end{bmatrix} \end{aligned} \quad (17)$$

Pole-face Rotation Matrix The first-order  $\mathbf{R}$  matrix for a pole-face rotation as shown in Fig.3 is  $\mathbf{R}(\beta) =$

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ h \tan \beta & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & -h \tan(\beta - \psi) & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad (18)$$

where  $\beta = \text{angle of rotation of pole}$ ,  $1/h = \rho_0 = \text{bending radius of central orbit}$ ,  $g = \text{total gap of the dipole magnet}$ ,  $\psi = \text{a correction term resulting from the finite extent of the fringing fields}$ ,

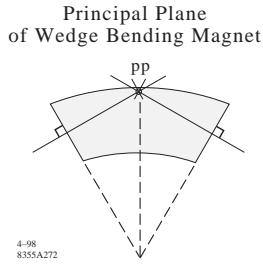


Figure 4: Uniform wedge bend whose two principal planes meet at the center of magnet.

$\psi = k(hg) \frac{1 + \sin^2 \beta}{\cos \beta}$ ,  $k$  is typically of the order  $1/2$  to  $1$ , so if  $hg$  is less than  $1\%$ , this correction term can usually be ignored.

Note that a pole-face rotation on a bending magnet is equivalent to a thin quad located at the input or exit face of the bending magnet having a focal length  $1/F_x = -h \tan \beta$  in the  $x$  plane and  $1/F_y = +h \tan \beta$  in the  $y$  plane.

For a bending magnet having rotated input and output pole faces, the total  $\mathbf{R}$  Matrix for the magnet is

$$\mathbf{R} = \mathbf{R}(\beta_2) \mathbf{R}_{\text{wedge bend}} \mathbf{R}(\beta_1) \quad (19)$$

Solenoid (see also Sec.4.7.4)  $\mathbf{R}_{\text{sol}} =$

$$\begin{bmatrix} C^2 & \frac{SC}{K} & SC & \frac{S^2}{K} & 0 & 0 \\ -KSC & \frac{C^2}{K} & -KS^2 & SC & 0 & 0 \\ -SC & -\frac{S^2}{K} & C^2 & \frac{SC}{K} & 0 & 0 \\ KS^2 & -SC & -KSC & C^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad (20)$$

where  $L = \text{effective length of solenoid}$ ,  $K = B_0/(2B\rho)$ ,  $B_0$  is the field inside the solenoid,  $(B\rho)$  is the momentum of the central trajectory,  $C = \cos KL$ ,  $S = \sin KL$ .

Rotating the transverse coordinates  $x$  and  $y$  about the optical axis at the exit of the solenoid by an angle  $(-KL)$  decouples the  $x$  and  $y$  first-order terms, i.e.  $\mathbf{R}_{\text{rot}}(-KL) \mathbf{R}_{\text{sol}} =$

$$\begin{bmatrix} C & S/K & 0 & 0 & 0 & 0 \\ -KS & C & 0 & 0 & 0 & 0 \\ 0 & 0 & C & S/K & 0 & 0 \\ 0 & 0 & -KS & C & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad (21)$$

The focal length of a solenoid is  $1/F_x = 1/F_y = -R_{21} = -R_{43} = KS = K \sin KL$ .

Beam rotation matrix

$$\mathbf{R}_{\text{rot}} = \begin{bmatrix} C & 0 & S & 0 & 0 & 0 \\ 0 & C & 0 & S & 0 & 0 \\ -S & 0 & C & 0 & 0 & 0 \\ 0 & -S & 0 & C & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad (22)$$

where  $\alpha$  = the angle of rotation about the  $z$  axis (the direction the beam is going),  $C = \cos \alpha$ ,  $S = \sin \alpha$ . For example a quadrupole rotated clockwise by an angle  $\alpha$  about the  $z$  axis is

$$\mathbf{R}_{\text{rot}}(-\alpha) \mathbf{R}_{\text{quad}} \mathbf{R}_{\text{rot}}(+\alpha) = \mathbf{R}_{\text{rot, quad}} \quad (23)$$

A skew quad is generated by setting  $\alpha = \pm 45^\circ$ .

If the convention is that  $\mathbf{R}_{\text{bend}}$  represents a magnet bending to the right, then a bend magnet bending the beam up is given by the coordinate rotation

$$\text{BEND UP} = \mathbf{R}(+\alpha) \mathbf{R}_{\text{bend}} \mathbf{R}(-\alpha) \quad (24)$$

$$\text{BEND DOWN} = \mathbf{R}(-\alpha) \mathbf{R}_{\text{bend}} \mathbf{R}(+\alpha) \quad (25)$$

where  $\alpha = 90^\circ$ . The case  $\alpha = 180^\circ$  represents a magnet bending to the left in either of the above equations.

**Accelerator matrix** The transport matrix for an accelerator section is listed in [3].

**References**

- [1] K. Brown et al, SLAC-75-rev-4 (1982); SLAC-91-rev-2 (1977)
- [2] K. Brown, R. Servranckx, SLAC-PUB-3381 (1984)
- [3] D.C. Carey, K.L. Brown, F. Rothacker, SLAC-R/95/462 (1995) p.179-182

## 2.2.2 3-D Multipole Expansion, Calculation of Transfer Maps from Field Data, Fringe Fields

M. Venturini, LBNL  
A. Dragt, U. Maryland

The behavior of orbits in charged-particle beam transport and accelerator systems can depend sensitively on nonlinear fringe-field and high-order-multipole effects in the various beam-line elements. Calculation of transfer maps that capture these effects requires a detailed and realistic model of the interior and fringe fields, including their high spatial derivatives. We present methods to determine field representations from available field data that are suitable for this purpose.

**Cylindrical multipole analysis** In a current-free region the magnetic field  $\mathbf{B} = \nabla\psi$  can be described in terms of a magnetic scalar potential  $\psi$  obeying  $\nabla^2\psi = 0$ . For straight beam-line elements (solenoids, quadrupoles, higher-order multipole magnets, wigglers, and dipoles with small sagitta) a cylindrical system with the  $z$ -axis parallel to the magnet axis represents a natural coordinate choice yielding for  $\psi$  the representation  $\psi = \sum_{m=0}^{\infty} \psi_{m,c}(\rho, z) \cos m\phi + \sum_{m=1}^{\infty} \psi_{m,s}(\rho, z) \sin m\phi$ , with  $\psi_{m,s} = \int dk e^{ikz} b_m(k) I_m(k\rho)$  and  $\psi_{m,c}$  being the same as  $\psi_{m,s}$  but with  $a_m(k)$  replacing  $b_m(k)$ . The quantities  $I_m$  are the modified Bessel functions. The arbitrary functions  $b_m$  and  $a_m$  measure the amplitude of the ‘normal’ and ‘skew’ field components. Purely solenoidal, dipole, quadrupole, sextupole, ... field components correspond to  $m = 0, 1, 2, 3, \dots$ , respectively. Equivalently we can write  $\psi_{m,\alpha}(\rho, z) = \sum_{\ell=0}^{\infty} d_{\ell,m} G_m^{[2\ell]}(z) \rho^{2\ell+m}$ , where  $d_{\ell,m} = (-1)^\ell / [2^{2\ell} \ell!(\ell+m)!]$ , and  $\alpha = s, c$  [1, 2, 3]. The functions  $G_{m,\alpha}^{[n]}(z)$ , called the on-axis generalized gradients, obey  $G_{m,\alpha}^{[n+1]}(z) = \frac{d}{dz} G_{m,\alpha}^{[n]}(z)$ , and uniquely determine  $\psi$ . The resulting magnetic field for a normal pure multipole of order  $m \geq 1$  reads

$$\begin{aligned} B_\rho &= \sin m\phi \sum_{\ell=0}^{\infty} (2\ell + m) d_{\ell,m} G_m^{[2\ell]}(z) \rho^{2\ell+m-1} \\ B_\phi &= \cos m\phi \sum_{\ell=0}^{\infty} m d_{\ell,m} G_m^{[2\ell]}(z) \rho^{2\ell+m-1} \\ B_z &= \sin m\phi \sum_{\ell=0}^{\infty} d_{\ell,m} G_m^{[2\ell+1]}(z) \rho^{2\ell+m} \end{aligned}$$

The expressions for the skew multipoles follow from the replacements  $\sin \rightarrow \cos$  and  $\cos \rightarrow -\sin$ . For pure solenoids  $B_z = \sum_{\ell=0}^{\infty} d_{\ell,0} G_0^{[2\ell+1]}(z) \rho^{2\ell}$ ,  $B_\rho = \sum_{\ell=1}^{\infty} 2\ell d_{\ell,0} G_0^{[2\ell]}(z) \rho^{2\ell-1}$ ,  $B_\phi = 0$ .

In a few cases, certain iron-free or rare-earth-cobalt (REC) magnets [3, 4, 5], it is possible to determine the generalized gradients directly in analytic form or in terms of integrals over current or REC material distributions. When iron is present, the problem is more difficult. Then usually the available information is in the form of field data at various discrete points obtained either by measurement or the use of 3-D finite element codes. In the following we discuss how to determine the generalized gradients from field data.

**Fitting methods** A common past approach has been to use on-axis or mid-plane 3-D field data to fit some kind of profile function specified in analytic form with various free parameters. Examples of fitting profiles are the Enge functions [6], of the form  $1/(1 + e^{S_k(z)})$ , with  $S_k(z)$  being a polynomial of order  $k$  in  $z$ . Other examples include the following model functions (corresponding to the exact fields generated by certain cylindrical sheet currents [3, 4]). For solenoids  $G_0^{[1]}(z) = g[F_0(z^+) - F_0(z^-)]$ , dipoles  $G_1^{[0]}(z) = g[F_1(z^+) - F_1(z^-)]$ , and quadrupoles  $G_2^{[0]}(z) = g[F_2(z^+) - F_2(z^-)]$ ; where  $F_0(t) = f_1(t)$ ,  $F_1(t) = 2f_1(t) - f_3(t)$ ,  $F_2(t) = 9f_1(t) - 8f_3(t) + 3f_5(t)$ , with  $f_n(t) = t^n/(a^2 + t^2)^{n/2}$  and  $z^\pm = z \pm L/2$ . The fitting parameters are the field/gradient strength  $g$  and the radius  $a$  and length  $L$  of the cylindrical sheet current. Once the parameters have been determined the profile functions are repeatedly differentiated as needed for map computation. However, differentiation amplifies the unavoidable errors present either because of imperfect fitting or noise in the field data. Use of these methods is therefore not expected to be reliable for calculations beyond first or perhaps second order.

**Surface methods (circular cylinder)** The problem of noise and its amplification by differentiation can be overcome with the use of surface methods. Suppose that the normal field component  $B_\rho(R, \phi, z)$  is known on the surface  $\rho = R$  of a virtual circular cylinder (in practice, the values of the field on the surface are obtained by interpolation of the field data on nearby grid points). The cylinder, long enough to include the fringe regions, is to be contained within the aperture of the magnetic device and aligned with the  $z$ -axis. This surface information is sufficient to uniquely determine the full magnetic field within the surface. From  $B_\rho(R, \phi, z)$  we compute the quantities  $\hat{B}_{\rho, m, s}(k) = \int_{-\infty}^{\infty} dz e^{-ikz} \int_0^{2\pi} d\phi \sin(m\phi) B_\rho(R, \phi, z)/(2\pi)^2$  with a similar expression for  $\hat{B}_{\rho, m, c}(k)$  with sin replaced by cos. Then the  $G_{m, \alpha}^{[n]}(z)$  are given in terms of surface data by the relations

$$G_{m, \alpha}^{[n]}(z) = \frac{i^n}{2^m} \int_{-\infty}^{\infty} dk \frac{k^{n+m-1}}{I_m'(kR)} \hat{B}_{\rho, m, \alpha}(k) e^{ikz} \quad (1)$$

The integrand in (1) decays rapidly at large  $|kR|$  because of the Bessel function property

$|1/I_m'(x)| \sim \sqrt{2\pi|x|}e^{-|x|}$  for  $|x| \gg 1$ . This rapid decay provides a powerful intrinsic smoothing effect that attenuates any uncorrelated errors/noise present in the field data [1, 2, 7]. The smoothing increases with  $R$ , hence motivating the use of field data on the largest radius cylinder that can be accommodated within the aperture of the device. Expressions similar to (1) can be worked out for  $B_z$  or  $B_\phi$  and may be more useful depending on circumstances. Variants employ sets of orthonormal basis functions solving the Laplace equation when cylinders with elliptic or rectangular cross sections are used. Their use improves the smoothing property of the method for modelling magnets (e.g. wigglers) with an aperture that is significantly larger in one direction [2, 7]. Analogous methods are applicable to rf cavities [8].

**More general surfaces** In the case of curved elements with large sagitta, it is necessary to use more general surfaces (e.g. a bent box with straight ends). For geometries in which the Laplace equation is not separable, it is possible to obtain explicit expressions for interior fields in terms of surface fields providing that all *three* components of the surface field are employed (vs. *one* field component required for geometries in which the Laplace equation is separable, as in the previous discussion). One method is based on the Helmholtz decomposition theorem for a source-free vector field in a volume  $V$  bounded by a surface  $S$  [9, 10]. Let  $\hat{n}(\mathbf{r}')$  be the outward unit normal to  $S$  at the point  $\mathbf{r}' \in S$ . In this decomposition  $\mathbf{B}(\mathbf{r}) = \nabla\chi + \nabla \times \mathbf{A}$  with  $\chi(\mathbf{r}) = (4\pi)^{-1} \int_S dS' [\hat{n}(\mathbf{r}') \cdot \mathbf{B}(\mathbf{r}')]/|\mathbf{r} - \mathbf{r}'|$  and  $\mathbf{A}(\mathbf{r}) = -(4\pi)^{-1} \int_S dS' [\hat{n}(\mathbf{r}') \times \mathbf{B}(\mathbf{r}')]/|\mathbf{r} - \mathbf{r}'|$ . A second method, which employs *only* a vector potential so that  $\mathbf{B}(\mathbf{r}) = \nabla \times \mathbf{A}$ , utilizes the Dirac monopole vector potential and is suitable for a Hamiltonian formulation [2, 7, 11]. It is based on the decomposition  $\mathbf{A} = \mathbf{A}^n + \mathbf{A}^t$  where  $\mathbf{A}^n(\mathbf{r}) = \int_S dS' [\hat{n}(\mathbf{r}') \cdot \mathbf{B}(\mathbf{r}')]\mathbf{K}^n(\mathbf{r}, \mathbf{r}')$  and  $\mathbf{A}^t(\mathbf{r}) = \int_S dS' \psi(\mathbf{r}') \mathbf{K}^t(\mathbf{r}, \mathbf{r}')$ . Here  $\psi(\mathbf{r}')$  is the scalar potential on  $S$ , and the kernels  $\mathbf{K}^n$  and  $\mathbf{K}^t$  are given by the relations  $\mathbf{K}^n(\mathbf{r}, \mathbf{r}') = \{\hat{n}(\mathbf{r}') \times (\mathbf{r} - \mathbf{r}')\}/\{4\pi|\mathbf{r} - \mathbf{r}'|[(\mathbf{r} - \mathbf{r}') \cdot \hat{n}(\mathbf{r}')]\}$  and  $\mathbf{K}^t(\mathbf{r}, \mathbf{r}') = [\hat{n}(\mathbf{r}') \times (\mathbf{r} - \mathbf{r}')]/[4\pi|\mathbf{r} - \mathbf{r}'|^3]$ . Both methods are intrinsically smoothing and yield analytic interior fields. The second yields interior fields that satisfy the Maxwell equations exactly regardless of field data quality and integration accuracy.

**Hard-edge fringe-field maps** The accurate treatment of high-order fringe-field effects is complicated. They can be computed reliably if interior fields are known in analytic form, as is the case for iron-free or REC magnets, or if surface methods are used. In some few cases, hard-edge models provide a useful first approximation. Suppose the transfer map for a magnet is written in the factorized Lie form  $\mathcal{M} = \mathcal{R} \exp(.f_3:) \exp(.f_4:) \cdots$  (Sec.2.3.6) and that some model is proposed for the fringe-field transition from no field outside to some  $s$ -independent field inside. With such a model one may study the factors of  $\mathcal{M}$  in the limit in which the transition region shrinks to zero. Do the various factors then approach definite finite limits, and are the limits model independent? It can be shown that, for solenoids,  $\mathcal{R}$  and  $f_3$  do take on limiting values with these desirable properties. However, the  $f_4$  and higher-order polynomials become infinite. In the case of dipoles,  $\mathcal{R}$  and  $f_3$  also take on finite and model-independent limiting values. However, some of the  $f_4$  take on finite model independent values, some take on finite but model dependent values, and some become infinite. Thus, in the case of solenoids and dipoles, there is no meaningful hard-edge fringe-field approximation beyond second order; detailed knowledge of the field profile is essential to go beyond second order. For the case of quadrupoles it can be shown that  $\mathcal{R}$ ,  $f_3$ , and  $f_4$  take on finite and model-independent limiting values. Thus, for quadrupoles, there is a meaningful hard-edge approximation through third order. The limiting behavior of still higher-order multipoles is also known [12]. However, use of such information would seem warranted only if all other relevant beam-line elements are treated accurately, including their fringe-field effects, through the same order.

In the dipole case the limiting form of  $\mathcal{R}$  is built into most codes and is referred to as *edge focusing* [12]. The hard-edge form of  $f_3$  for dipoles is available as an option in MaryLie [5]. No detailed studies have been carried out to see how well the hard-edge approximation holds in the case of realistic dipoles, but such studies should now be possible with the use of surface methods.

The hard-edge form of  $f_4$  for quadrupoles is given by  $f_4^\pm = \pm(k/12)(y^3 p_y - x^3 p_x + 3x^2 y p_y - 3y^2 x p_x)$ , and its use is equivalent to Lee-Whiting's earlier result [12, 13]. Here  $k = Q/B\rho$  where  $Q$  is the quadrupole strength,  $B\rho$  is

the design-orbit rigidity, and  $\pm$  refers to entrance and exit. Numerical evaluation of the map for a REC quadrupole in the limit of zero inner radius verifies that the exact  $f_4$  does indeed approach the hard edge limit, and it does so fairly rapidly. This example shows that, for quadrupoles, there are cases where the hard-edge approximation can be usefully employed even when the aperture is quite large.

## References

- [1] M. Venturini, A. Dragt, NIM-A 427, 387 (1999)
- [2] A. Dragt, Lie Methods for Nonlinear Dynamics with Applications to Accelerator Physics (2010); [www.physics.umd.edu/dsat/](http://www.physics.umd.edu/dsat/)
- [3] M. Bassetti, C. Biscari, PA 52, 221 (1996)
- [4] P. Walstrom, NIM-A 519, 216 (2004)
- [5] A. Dragt et al., MaryLie Manual (2003); [www.physics.umd.edu/dsat/](http://www.physics.umd.edu/dsat/)
- [6] H.A. Enge, Rev. Sci. Instrum. 35, 278 (1964)
- [7] C. Mitchell, A. Dragt, PRST-AB 13, 064001 (2010)
- [8] D. Abell, PRST-AB 9, 052001 (2006)
- [9] R. Plonsey, R.E. Collin, Principles and Applications of Electromagnetic Fields, McGraw-Hill (1961)
- [10] S. Manikonda, M. Berz, NIM-A 558, 175 (2006)
- [11] A. Dragt et al, arXiv:1012.1647 (2000)
- [12] E. Forest, Beam Dynamics, Harwood (1998)
- [13] G. Lee-Whiting, NIM 83, 232 (1972)

### 2.2.3 Lattices for Collider Storage Rings E. Keil, CERN

Symbols  $\alpha, \beta, \gamma$  Courant-Snyder parameters,  $K = (1/B\rho)(dB_y/dx)$ ,  $f$  focal length of quadrupoles,  $L_p$  Length of whole cell,  $\mu$  Phase advance per cell,  $\varphi$  Bending angle of whole cell,  $z$  Transverse coordinate  $x$  or  $y$ .

**Lattice matching** A lattice typically is composed of unit cells and insertions. The points where cells and insertions meet require lattice matching.

Insertions transport the beam from a point with one set of  $\alpha, \beta, D, D'$  to another point with a different set of these parameters. Phase advances through the insertion may also be constrained. Splitting an insertion into modules helps, with each module achieving part of the matching.

Matching is done by matching programs (Sec.1.7), which adjust quadrupole and dipole

strengths, drift space lengths. The number of variables  $\geq$  that of conditions.

**Separated-function FODO cell** Two quads, with equal and opposite  $f$ , assumed to have vanishing length, separated by length  $L_p/2$ . Spaces between quadrupoles are filled by homogeneous-field dipoles.  $L_p$  applies to both focusing and bending. Assume that edge focusing in dipoles is negligible (when  $\rho \gg L_p$ ) and that  $\varphi \ll 1$ . In terms of  $\mu$  (assumed the same in  $x$ - and  $y$ -planes) and  $L_p$  [1],

$$f = \pm \frac{L_p}{4 \sin \frac{\mu}{2}} \quad (1)$$

$$\beta^\pm = \frac{L_p(1 \pm \sin \frac{\mu}{2})}{\sin \mu} \quad (2)$$

$$\alpha^\pm = \frac{\mp 1 - \sin \frac{\mu}{2}}{\cos \frac{\mu}{2}} \quad (3)$$

$$D^\pm = \frac{L_p \varphi [1 \pm \frac{1}{2} \sin \frac{\mu}{2}]}{4 \sin^2 \frac{\mu}{2}} \quad (4)$$

The sign of  $\alpha$  applies to the quadrupole entrances. At the exits,  $\alpha$  has the opposite sign. The upper (lower) signs apply to the focusing (defocusing) quadrupole in the plane under consideration. From the focusing quadrupole ( $s = 0$ ) to the defocusing quadrupole ( $s = L_p/2$ ),

$$\beta(s) = \beta^+ - \frac{2s(1+\sin \frac{\mu}{2})}{\cos \frac{\mu}{2}} + \frac{4s^2}{L_p} \tan \frac{\mu}{2} \quad (5)$$

$$D(s) = D^+ \left( 1 - \frac{2s \sin \frac{\mu}{2}}{L_p} \right) + \frac{s^2 \varphi}{2L_p} \quad (6)$$

See Figs.1 and 2.  $\beta^+/L_p$  has a minimum near  $\mu/2\pi = \sin^{-1}(\frac{\sqrt{5}-1}{2})/\pi \approx 0.21$ . The dispersion averaged over the FODO cell is

$$\langle D \rangle = \frac{L_p \varphi}{4} \left( \frac{1}{\sin^2 \frac{\mu}{2}} - \frac{1}{12} \right) \quad (7)$$

The momentum compaction factor is (see also Sec.2.3.11)

$$\alpha_c = \frac{\varphi \langle D \rangle}{L_p} = \left( \frac{\varphi}{2} \right)^2 \left( \frac{1}{\sin^2 \frac{\mu}{2}} - \frac{1}{12} \right) \quad (8)$$

The natural chromaticity per FODO cell is

$$\xi_{\text{FODO}} = -\frac{1}{\pi} \tan \frac{\mu}{2} \quad (9)$$

If a storage ring consists of  $n$  FODO cells, it has a total natural chromaticity  $\xi = n\xi_{\text{FODO}}$ . The synchrotron radiation integrals  $\mathcal{I}_5$  and  $\mathcal{I}_6$  (Sec.3.1.4)

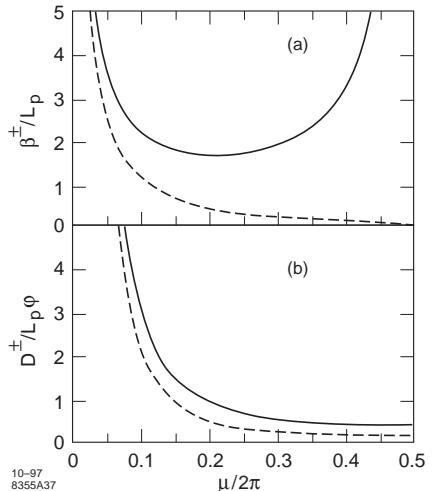


Figure 1: (a)  $\beta^\pm/L_p$  vs  $\mu/2\pi$ , (b)  $D^\pm/L_p\varphi$  vs  $\mu/2\pi$ .

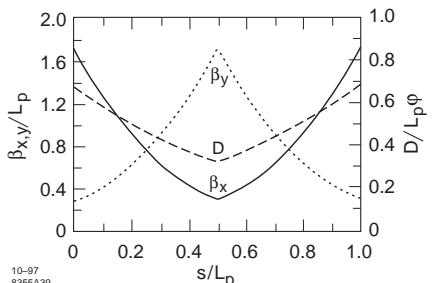


Figure 2:  $\beta_x/L_p$ ,  $\beta_y/L_p$ , and  $D/L_p\varphi$  vs  $s/L_p$  for the FODO case  $\mu/2\pi = 0.25$ .

per cell are

$$\mathcal{I}_{5,\text{FODO}} = \frac{\varphi^5 (1 - \frac{3}{4} \sin^2 \frac{\mu}{2} + \frac{1}{60} \sin^4 \frac{\mu}{2})}{4L_p \sin^2 \frac{\mu}{2} \sin \mu} \quad (10)$$

$$\mathcal{I}_{6,\text{FODO}} = \frac{\varphi^2}{2\ell_Q} \left( 1 + \frac{4}{\sin^2 \frac{\mu}{2}} \right) \quad (11)$$

where  $\ell_Q$  = quad length.

In a flat ring, the dispersion function often scales approximately with  $\sqrt{\beta}$ ,

$$D(s) \approx \sqrt{\frac{\alpha_c R}{\nu_x}} \sqrt{\beta_x(s)} \quad (12)$$

where  $2\pi R$  is the ring circumference and  $\nu_x$  is the horizontal tune of the ring.

The average  $\beta$ - and dispersion functions are given approximately by, using Eqs.(2) and (4) for

relatively small  $\mu$ ,

$$\langle \beta_{x,y} \rangle \approx R/\nu_{x,y}, \quad \langle D \rangle \approx R/\nu_x^2 \quad (13)$$

where  $2\pi R = nL_p$ , and  $\nu_{x,y} = n\mu_{x,y}/2\pi$ . It follows from Eq.(13) that

$$\alpha_c \approx 1/\nu_x^2, \quad \gamma_t \approx \nu_x \quad (14)$$

(For nonlinear momentum compaction factor for a FODO lattice, see Sec.2.3.11.)

**Combined-function FD cell** Two combined-function dipoles of length  $L_p/2$  each, with equal and opposite strengths  $K$ , and no drift spaces.

$$\begin{aligned} \cos \mu &= \cos \frac{L_p \sqrt{K}}{2} \cosh \frac{L_p \sqrt{K}}{2} \\ &= 1 - \frac{L_p^4 K^2}{96} + \frac{L_p^8 K^4}{645120} + \dots \end{aligned} \quad (15)$$

Equating the leading two terms in Eq.(15) to the leading two terms of  $\cos \mu$  yields (overestimating  $K$  by <10% for phase advances  $\mu/2\pi \leq 1/4$ ) [2]

$$K \approx \frac{4\mu\sqrt{3}}{L_p^2} \quad (16)$$

$\beta^\pm$  at the centers of the F and D quadrupoles are (<10% error for  $\mu/2\pi \leq 1/4$ )

$$\beta^\pm \approx \frac{L_p}{\sin \mu} \left( 1 \pm \frac{\pi\sqrt{3}}{2} \frac{\mu}{2\pi} \right) \quad (17)$$

while (underestimating  $D^\pm$  by <13% for  $\mu/2\pi \leq 1/8$ )

$$D^\pm \approx \frac{L_p \varphi}{\mu^2} \left( 1 \pm \frac{\mu\sqrt{3}}{8} \right) \quad (18)$$

See Fig.3.

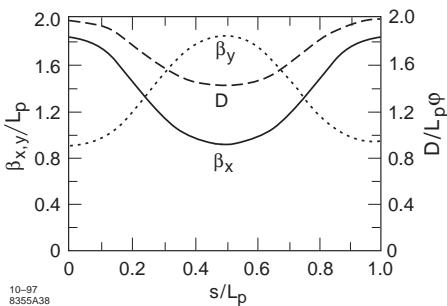


Figure 3: Computed (no Taylor series approximation)  $\beta_x/L_p$ ,  $\beta_y/L_p$  and  $D/L_p\varphi$  vs  $s/L_p$  for the combined FD case  $\mu/2\pi = 0.125$ .

**Dispersion suppressor insertion** These insertions match  $D$  without affecting  $\alpha, \beta$ . The idea is to bring  $D$  from the value in the FODO cells to zero by launching a forced oscillation of  $D$  around a value  $\sim \frac{1}{2}$  of that in the FODO cells. (i) If  $\pi/\mu = \text{integer}$ ,  $D$  is suppressed by  $\pi/\mu$  FODO cells with bending angle  $\varphi/2$ , using Eq.(4). (ii) It is possible to suppress  $D$  within two FODO cells, if their bending angles are chosen as [3]

$$\varphi_1 = \varphi \left( 1 - \frac{1}{4 \sin^2 \frac{\mu}{2}} \right), \quad \varphi_2 = \frac{\varphi}{4 \sin^2 \frac{\mu}{2}} \quad (19)$$

where  $\varphi_1$  and  $\varphi_2$  are bending angles of the FODO cell on the side where  $D \neq 0$ , and  $D = 0$ , respectively. The betatron oscillations are unchanged in both styles of dispersion suppressor. When  $\mu = \pi/2$ , both styles coincide. For  $\mu \leq \pi/3$ , there are no reversed dipoles in the dispersion suppressor. Both styles can start at horizontally focusing or defocusing quadrupoles [4]. However, they both work for just one phase advance. (iii) It is possible to suppress  $D$  by just modifying the focal length of two quadrupoles. In this case,  $\alpha, \beta$  and  $\mu$  are also modified. (iv) In hybrid dispersion suppressors, some bending magnets are left out, and  $D$  is halved by halving the product  $L_p \varphi$ , using Eq.(4). By giving them enough free parameters, they can be matched to FODO cells over a range of phase advances [5].

**Low-beta insertion** These insertions reduce the  $\beta$ -functions to small values, assuming  $D = 0$  already. They are used in colliders to achieve small beam sizes at the IPs. Lattice matching requires four or six free parameters, typically quadrupole gradients, depending on whether only two  $\alpha$ 's and two  $\beta$ 's, or whether also two  $\mu$ 's are matched. Usually  $\alpha_x = \alpha_y = D = D' = 0$  at the IP. Most critical are the quadrupole doublet or triplet near IP.

The  $\beta$ -function near a waist is given by

$$\beta(s) = \beta_0 + (s - s_0)^2 / \beta_0 \quad (20)$$

where  $\beta_0$  is the  $\beta$ -function at waist position  $s_0$ .

Doublets are often used in  $e^+e^-$  colliders with flat beams, the first quadrupole from the IP focuses vertically, the second horizontally. In thin lens approximation, their focal lengths  $d_1$  and  $d_2$  are found by imposing point-to-parallel focusing in both planes,

$$d_1 = \pm \frac{s_1 \sqrt{s_2}}{\sqrt{s_1+s_2}}, \quad d_2 = \mp \sqrt{s_2} \sqrt{s_1+s_2} \quad (21)$$

where  $s_{1,2}$  are the drift spaces from the IP to the first quadrupole, and between the two quadrupoles, respectively. The two solutions correspond to interchanging  $x$  and  $y$ . In LEP, the point-to-parallel approximation overestimates the focal length by  $< 8\%$ , and the thin-lens approximation by another  $12\%$ .

Triplets are often used in pp and  $p\bar{p}$  colliders with round beams. Assuming a symmetrical triplet with the first and third quadrupole having the same strength and equi-distant from the central quadrupole, and in thin lens approximation, the focal lengths  $t_1$  and  $t_2$  are found by imposing point-to-parallel focusing in both planes,

$$\begin{aligned} t_1 &= \pm \sqrt{\frac{-r + s_2 (2 s_1^2 + 3 s_1 s_2 + 2 s_2^2)}{2(s_1 + s_2)}} \\ t_2 &= \frac{r - s_2 (2 s_1^2 + 5 s_1 s_2 + 2 s_2^2)}{4 (s_1 + s_2) t_1} \quad (22) \\ r &= [s_2^2(2s_1^2+3s_1s_2+2s_2^2)^2-8s_1^2s_2^3(s_1+s_2)]^{\frac{1}{2}} \end{aligned}$$

where  $s_{1,2}$  are the drift spaces from the IP to the first quadrupole, and between the quadrupoles, respectively. In LHC (Version 4.2) [6], the point-to-parallel approximation overestimates  $t_i$  by  $< 8\%$ , and the thin-lens approximation by another  $< 36\%$ .

**Two-ring Colliders** Two rings are needed in colliders of particles with the same sign of charge, e.g. LHC and RHIC, and in colliders of particles with opposite charge, but different momentum. Since luminosity (Sec.2.5)  $\mathcal{L}$  is proportional to beam current  $I$ , two rings are also needed in “factories” of particles with opposite charge and high  $\mathcal{L}$ , when  $I \gg$  bunch current, limited by collective effects (Sec.2.4). The rings must have equal circumferences [7]. Arcs often consist of FODO cells and dispersion suppressors, described above, and sometimes of separated function lattices with bending period length a multiple of focusing period length. Two-ring interaction regions consists of low- $\beta$  insertions, described above, for each beam, possibly with common elements, and separation schemes that funnel the two beams from separate arcs through the interaction point IP with low  $\beta$ -values and back into separate arcs. Beams often cross at an angle  $2\theta$ , often measured in units of Piwinski parameter  $\Phi = \theta\sigma_s/\sigma_\perp$  (Sec.2.3.4) with rms bunchlength  $\sigma_s$  and rms radius  $\sigma_\perp$  in crossing plane. Large crossing angle means

$\Phi > 1$ . Synchrotron radiation from magnets near IP should not get into detector in two-ring colliders with at least one  $e^\pm$  beam. Cf. [8, 9] for reviews of high-luminosity  $e^+e^-$  colliders.

Separation schemes (Sec.2.5.4.1) depend on whether the beams have equal or opposite charge and equal or different momenta. In a pp collider like LHC [10], beams are in a common vacuum chamber, and focused by common quadrupoles between separating dipoles on either side of IP. In  $e^+e^-$  B factories like KEK-B and PEP-II, and ep colliders like HERA [11], different momenta of beams are used to achieve separation. In  $e^+e^-$  colliders like BEPC-II [12] and DAΦNE with beams of opposite charge and equal momentum, the beams cross horizontally at an angle (Sec.2.5.2). The crossing angle between the beams is enhanced by the common first horizontally defocusing quadrupole, such that the second horizontally focusing quadrupoles can be installed side-by-side.

Crossing angles  $\theta$  are used in almost all two-ring colliders to avoid head-on beam-beam collisions (Sec.2.5.4.1) other than those at the IP. Long-range beam-beam collisions still occur in the common vacuum chamber at every half bunch spacing near the IP. The larger  $\theta$ , the smaller the long-range kicks and tune shifts. In LHC [10] crossings are vertical and horizontal in diametrically opposite IPs. Separation at long-range collisions is about  $7\sigma_\perp$ . See also Sec.2.5.4.

Bunch overlap at  $\Phi > 0$ , improved by crab crossings and crab waists, increases  $\mathcal{L}$ . In a crab crossing [13], used in KEK-B [14], a transversely deflecting RF cavity in each beam causes an orbit distortion that varies along the bunch. Hence, the bunches have a correlation at the IP between offset in the crossing plane and longitudinal position along the bunch such that  $\theta$  is cancelled. The crab waist [15] in DAΦNE (Sec.2.5.2) uses the fact that a test particle with an offset in the plane of crossing encounters the centre of the opposite bunch at a distance from the IP along the beam that is proportional to the offset. Sextupoles satisfying conditions on phases and strengths are used to optimize the geometric overlap between the two colliding bunches, and more importantly, to mitigate synchrobetatron resonances driven by beam-beam collisions with  $\Phi \neq 0$  (cf. Sec.2.3.4).

## References

- [1] E. Keil, CERN 77-13 (1977) 11

- [2] H. Bruck, Accélérateurs circulaires de particules, Presses universitaires de France (1966)
- [3] E. Keil, CERN 77-13 (1977) 22
- [4] R.H. Helm, SLAC-PUB-3278 (1984)
- [5] LEP Design Report, Vol.II, CERN-LEP/84-01 (1984)
- [6] The Large Hadron Collider: Conceptual Design, CERN/AC/95-05(LHC) (1995)
- [7] K. Hirata, E. Keil, Phys. Lett. B232 (1989) 413
- [8] Y. Funakoshi (ed.), ICFA Beam Dynamics Newsletter No.31 (2003)
- [9] M.E. Biagini (ed.), ICFA Beam Dynamics Newsletter No.48 (2009)
- [10] LHC Design Report, Vol.1, CERN-2004-003 (2004) <http://lhc.web.cern.ch/lhc/LHCDesignReport.html>
- [11] M. Seidel, F. Willeke, EPAC00 (2000) 379
- [12] J.Q. Wang et al, PAC07 (2007) 53
- [13] R.B. Palmer, Snowmass Conf. High Energy Physics in the 1990s, 1988, Ed. S. Jensen, World Scientific (1989) 642
- [14] Y. Funakoshi et al, EPAC08 (2008) 1893
- [15] <http://arxiv.org/ftp/physics/papers/0702/0702033.pdf>

## 2.2.4 Lattices for Low-Emissittance Light Sources

*C. Steier; LBNL*

Third generation light sources have been extremely popular since the first ones went into operation around 1992. They are based on lepton storage rings, in most cases using electron beams, in very few cases positron beams. The fundamental properties of these facilities are high beam currents ( $\geq 100$  mA), small emittances ( $\leq 10$  nm horiz.), moderate beam energy spreads ( $\leq 10^{-3}$ ) and the use of insertion devices (mostly undulators) in long ( $\sim 5$  m) straight sections to produce high brightness (Sec.4.2) photon beams. Originally there were two main lattice types that were optimized to meet these requirements: the double bend achromat (DBA [1]), first used in the NSLS at BNL, and the triple bend achromat (TBA [2]), first developed for the ALS at LBNL. Later on, facilities evolved further by detuning the achromat lattices to allow dispersion leakage into the straight sections. Newer light sources also use multi-bend achromats, as well as damping wigglers to reduce the emittance further.

### 2.2.4.1 Lattice choices

Traditionally the straights were designed with zero dispersion, which minimizes synchrobeta-

tron coupling, avoids beam size increase due to energy spread and results in the largest possible reduction of the equilibrium emittance due to radiation emitted in insertion devices. All early 3rd generation light sources employed achromat lattices. Over time, however, nearly all of the early 3rd generation rings moved away from the achromatic condition resulting in lower equilibrium emittances [3]. In most cases the dispersion in the straights is still small enough that insertion devices continue to reduce the overall equilibrium emittance and the effective emittance is smaller than in the achromatic case.

The wavelength  $\lambda$  of radiation emitted by an undulator is given by

$$\lambda = \frac{\lambda_u}{2n\gamma^2} \left( 1 + \frac{K^2}{2} \right) \quad (1)$$

where  $\lambda_u$  is the undulator period,  $\gamma$  is the relativistic Lorentz factor,  $K \approx \lambda_u B e / (2\pi m_0 c)$  (the “undulator parameter”),  $B$  is the magnetic field, and  $n = 1, 3, 5, \dots$  denotes the harmonic. The term  $\lambda_u / (2\gamma^2)$  shows that one can arrive at the same wavelength using higher beam energies and longer periods, or lower beam energies and shorter periods. Many parameters enter in the evaluation of the optimum beam energy, including cost, natural emittance, intrabeam scattering, beam instabilities, heat load on optics, magnetic material properties, and the desired photon wavelength range. In general, even with advanced undulator technology, low energy rings ( $\sim 2$  GeV) are unchallenged at low photon energies (below 20 eV) and provide excellent performance up to 4 keV, intermediate energy rings ( $\sim 3$  GeV) from a few 100 eV to above 10 keV, and, if higher photon energies are needed, higher electron energies are necessary ( $\sim 4.5 - 7$  GeV).

**Brightness and coherence** The spectral photon brightness,  $B(\lambda)$  (Sec.3.1.5), of light emitted at wavelength  $\lambda$  from an insertion device in a synchrotron light source is

$$B(\lambda) = \frac{F(\lambda)}{(2\pi)^2 \sigma_{Tx} \sigma_{Ty} \sigma_{Tx'} \sigma_{Ty'}} \quad (2)$$

where  $F(\lambda)$  is the photon flux and  $\sigma_{Tx}, \sigma_{Ty}, \sigma_{Tx'}$ , and  $\sigma_{Ty'}$  are the wavelength dependent convolutions of the respective electron and photon beam sizes and divergences. For a given flux, maximum brightness is achieved when the electron beam emittance is reduced toward and beyond the intrinsic diffraction-limited emittance  $\lambda/(4\pi)$  of the

photons and when the electron beam beta function is close to the equivalent beta function of the diffraction ellipse.

Coherence is a measure of the degree to which the radiation can exhibit interference patterns. The fraction  $f_{coh}$  of photon flux at wavelength  $\lambda$  that is transversely coherent is related to the ratio of the intrinsic photon emittance to the total emittance of the photon beam,

$$f_{coh} = \frac{F_{coh,T}(\lambda)}{F(\lambda)} = \frac{\lambda/(4\pi)}{\sigma_{Tx}\sigma_{Tx'} \sigma_{Ty}\sigma_{Ty'}} \quad (3)$$

Flux, brightness, coherent flux, and coherent fraction are not the only figures of merit for a synchrotron light source, but they are always important considerations. Which of the four quantities is most important depends upon each particular experiment.

**Equilibrium emittance** The horizontal emittance  $\epsilon_x$  results from the combined effect of an excitation of horizontal betatron oscillations (Sec.3.1.4),  $S_x$ , and its damping,  $\tau_x$ ,

$$\epsilon_x = S_x \tau_x, \quad \frac{1}{\tau_x} = \frac{J_x C_\gamma E^3}{4\pi T_0} \oint \frac{1}{\rho^2} ds = \frac{J_x C_\gamma E^3}{4\pi T_0} I_2, \quad (4)$$

$$\epsilon_x = C_q \frac{\gamma^2 I_5}{J_x I_2}, \quad I_5 = \oint \frac{\gamma_x \eta_x^2 + 2\alpha_x \eta_x \eta_x' + \beta_x \eta_x'^2}{|\rho^3|} ds \quad (5)$$

with  $\rho$  the bending radius,  $J_x$  the horizontal damping partition number. The integrals along  $s$  are nonzero only in bending magnets and insertion devices, while they are zero in drift spaces and negligible in quadrupoles and sextupoles. In most machines the contributions of the insertion devices to  $S_x$  and  $\tau_x$  are negligible and the essential contribution comes from the bending magnets. A small emittance requires the use of a lattice with small  $\eta_x$  and  $\beta_x$  in the bending magnets. The first lattices to achieve this and to provide the space for insertion devices were the double and triple bend achromats [1, 2]. The principle is shown in Fig.1.

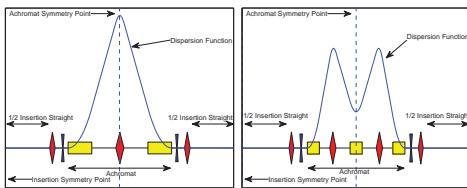


Figure 1: Principle of double and triple bend achromat lattices.

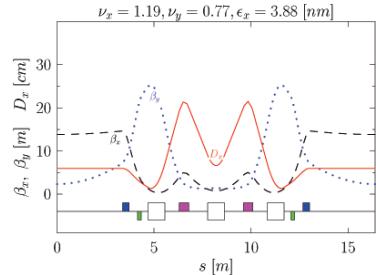


Figure 2: Example of a TBA lattice with distributed dispersion: ALS at LBNL.

The minimum emittance of a DBA lattice is given by

$$\epsilon_{DBA,min} = \frac{1}{4\sqrt{15}} \frac{C_q \gamma^2 \theta_b^3}{J_x} \quad (6)$$

where  $\theta_b$  is the bending angle per magnet. Lattices of facilities in operation are detuned significantly from these minimum values, to make optimization of nonlinear effects easier (see below). Instead of striving for minimum emittance in an achromatic condition, further optimization was achieved by allowing dispersion to leak into the straight sections, thereby lowering the dispersion in the arcs. One example of a distributed dispersion lattice is shown in Fig.2. The ALS lattice is a triple bend achromat which is operated with  $\eta_x = 6$  cm in the straights. The use of horizontally defocusing gradients in the dipole magnets allows for more compact lattices and also shifts the damping partition numbers, further reducing the horizontal emittance at the expense of the longitudinal damping.

How low one can get the natural emittance is related to the bending angle of individual magnets. Therefore newer rings with lower emittances generally have larger circumferences for a given beam energy, resulting in more unit cells. Alternatively, one can segment the bending into more magnets per unit cell and refocus both  $\beta_x$  and  $\eta$  between each pair of successive bending magnets. The resulting lattice is called a multiple bent achromat (MBA). For MBAs, the arcs usually resemble the so called theoretical minimum emittance structure [7]. The emittance from such a lattice scales asymptotically as

$$\epsilon_x \propto \frac{E^2}{N^3} \quad (7)$$

where  $N$  is the number of bending magnets. Fig.3 shows the lattice structure for a seven bend achromat lattice (Max-IV).

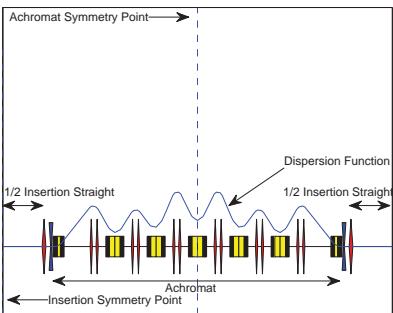


Figure 3: Example of a multi-bend achromat lattice, in this case a seven bend achromat.

#### 2.2.4.2 Chromaticity correction and nonlinear optimization

The low emittance design of light sources inherently relies on strongly focusing quadrupoles. These quadrupoles generate large chromatic aberrations that need to be corrected with sextupoles. The sextupoles in turn generate geometrical and nonlinear chromatic aberrations, exciting nonlinear resonances that can make the motion of the electrons unstable. This inherent feature has been the major design challenge from the beginning and continues to be a dominant topic for lattice upgrades as well as new rings. Many techniques have been developed to accurately predict the dynamic aperture (Sec.2.3.9) as well as the dynamic momentum aperture and to allow their optimization, thereby maximizing injection efficiency and Touschek lifetime (Sec.2.4.12). The techniques include design optimization tools, as well as measurement methods useful to optimize existing accelerators based on beam based measurements.

In terms of lattice designs, the trend has been to include more quadrupole families (the newest lattices include 6-10 quadrupole families) to allow full control of the linear lattice, and more sextupole families (state-of-the-art is 5-10 families) to allow control of detuning with amplitude terms as well as resonance driving terms, or a comprehensive optimization of the dynamic (momentum) aperture. Many rings have each magnet powered by an individual power supply to allow flexibility and improved correction of insertion device effects. The newest trends are the addition of weak octupole magnets to manipulate detuning with amplitude terms directly as well as the use of more relaxed lattices combined with damping wigglers.

Beyond correction of the chromaticity, while maintaining maximum dynamic and momentum aperture, lattice designers must also verify: (i) sensitivity of the dynamic (momentum) aperture to magnet imperfections or other lattice symmetry distortions; (ii) sensitivity to insertion devices as well as optimum compensation schemes; (iii) beam lifetime, in particular the influence of Touschek scattering; (iv) intrabeam scattering (Sec.2.4.12), which will be a very substantial effect in the latest light sources.

#### 2.2.4.3 Systematic lattice optimization techniques

Historically, lattice design depended strongly on the experience of the designer often choosing from a menu of known lattice choices. Many choices in this approach are subjective and usually the optimization of the nonlinear dynamics properties is carried out separately from the linear lattice design.

One newer technique to find the globally optimal lattices in terms of emittance or brightness is GLASS [4]. It uses a global grid scan of the few parameters of a simplified standard cell. It then analyzes the properties of all lattices that were stable. This allows to search for potential lattices with certain properties. Another technique involves multi-objective, genetic algorithms. Those algorithms are fairly old, but applications to accelerators are new. The first major one involved the optimization of a high brightness DC photoinjector. More recently they have been used for detailed linear lattice design and for multi-parameter optimization of the nonlinear dynamics of complex or low-periodicity lattices. They also allow for the simultaneous optimization of linear and nonlinear lattice properties [5].

These new techniques are now used for many purposes, including the optimization of special insertions (e.g. low beta straights, convergent beta straights, fs-slicing facilities, crab cavity insertions), development of lower emittance upgrades for existing facilities, optimization of linear and nonlinear lattices for new facilities, and the evaluation of candidate designs for ultimate storage rings.

#### 2.2.4.4 Evolution of light source lattices

3rd generation light sources have seen a steady evolution in the >15 years since their initial operation. This does not just include new and more

Table 1: Advanced storage ring facilities that are in early operation, under construction, or under study [8].

| Project   | Energy<br>[GeV] | Circ.<br>[km] | $\epsilon_x$<br>[nm] | Curr.<br>[A] | Lattice Design           | Status       |
|-----------|-----------------|---------------|----------------------|--------------|--------------------------|--------------|
| PETRA-III | 6               | 2.3           | 1                    | 0.1          | 7/8 FODO + 1/8 DBA + DW  | In operation |
| NSLS-II   | 3               | 0.792         | 0.6                  | 0.5          | 30 x DBA + DW            | Under const. |
| MAX IV    | 3               | 0.528         | 0.24                 | 0.5          | 20 x 7BA + DW            | Under const. |
| USRRLS    | 7               | 2             | 0.3                  | 0.5          | 50 x 4BA                 | Design study |
| XPS7      | 7               | 2.2           | 0.08                 | 1.0          | 80 x 6BA                 | Design study |
| PEP-X     | 4.5             | 2.2           | 0.1                  | 1.5          | DBA + TME + DW           | Design study |
| Tsumaki   | 6               | 1.44          | 0.07                 | 0.1          | 20x10BA, 4x5BA, 4xLSS+DW | Design study |
| USR7      | 7               | 3.1           | 0.015                | 0.2          | 40 x 10BA                | Design study |

advanced facilities, but also continuous improvements of existing ones. Lattice developments have evolved into two main directions. One is to allow for reduction of the natural emittance. The other is to incorporate complex local lattice insertions or breaks of the global lattice symmetry, to enable radiation sources with improved characteristics, while minimizing the negative effects on the global nonlinear beam dynamics. Some examples of those developments are listed below.

Advances to enable lower equilibrium emittances started with the move to distributed dispersion lattices. Later on, the addition of more sextupole families and most recently the inclusion of weak octupoles help to control the nonlinear dynamics and enabled lattices closer to the minimum emittance. Multi-bend achromats allow for smaller emittances for a given number of unit cells at the expense of a smaller fraction of available straight section space. More recently, damping wigglers have seen a resurgence for light sources. They allow to use a more relaxed lattice in the arc cells, resulting in larger momentum compaction factors and less lattice related nonlinear dynamics challenges while still achieving extremely small emittances. However, they use up space, require more expensive rf and cooling systems, increase the equilibrium energy spread and use up straight section space. Furthermore, nonlinear dynamics challenges due to wigglers can be significant. So to determine the optimum use of damping wigglers requires a careful design trade-off study. Tab.1 lists lattice parameters of the newest light sources in early operation or under construction, and of design studies for future facilities.

Examples of modifications to provide special photon sources include Superbends [6], a cost effective way to provide hard x-rays with moderate brightness using low beta function locations

in low energy rings. Other examples are the fs-slicing facilities, that use horizontal or vertical dispersion manipulation to provide spatial separation of energy modulated bunch slices to generate fs duration x-ray pulses. Intentional symmetry breaks were also introduced to allow for individual straight sections to be lengthened, enabling longer undulators, as well as for straight sections with smaller (or convergent) beta functions, to optimize the photon beam properties for specific beamlines. The newest proposal is a crab cavity insertion, which allows for two sets of crab cavities with precise control of phase advance and local coupling between them. Combined with proper beamline optics, this will allow to generate high brightness, high repetition rate x-ray pulses with ps duration.

#### 2.2.4.5 Ultimate storage rings

While storage rings are a “mature” technology, they nevertheless still have the potential for significantly further enhanced performance. One can imagine an “ultimate” storage ring [9, 10] that produces high-brightness, transversely coherent x-rays. For such a source to maximize transverse photon coherence, the beam emittance must be extremely small in both transverse planes, around the wavelength-dependent diffraction limit. Storage ring sources have achieved diffraction limited emittances for hard x-rays in the vertical plane by minimizing beam coupling, but the horizontal emittance must be reduced by a factor of 100 or more from the lowest values achieved today to reach that limit. Ultimate rings would necessarily have large circumferences (compare Tab.1) to reach low emittance values. Some of the main ideas are the use of more bending magnets, smaller physical apertures to allow for higher gradients of magnets, operation with large (or full) coupling to alleviate intrabeam scattering effects

and possible on-axis injection with beam replacement to mitigate small dynamic apertures. Ultimate rings would have brightness and coherent flux one or two orders of magnitude higher than the highest performance ring-based light sources in operation or under construction.

## References

- [1] M. Sommer, LAL/RT/83-15 (1983)
- [2] A. Jackson, PA 22 (1987) 111
- [3] L. Farvacque et al, EPAC 1994
- [4] D. Robin et al, PRST-AB 024002 (2008)
- [5] L. Yang et al, NIM A 609 (2009) 50; M. Borland, ICAP 2009
- [6] D. Robin et al, NIM A 538, 1-3 (2005) 65
- [7] S.Y. Lee, L. Teng, PAC 1991, p.2679
- [8] K. Balewski et al, PETRA III TDR, DESY 2004-035; J. Ablett et al, NSLS-II CDR, BNL 2006; M. Eriksson et al, EPAC 2008; A. Ropert et al, EPAC 2000; M. Borland, NIM A 557, 230 (2005); R. Hettel et al, PAC 2009; K. Tsumaki, N. Kumagai, NIM A 565, 394 (2006); M. Borland, Proc. SRI 2009
- [9] M. Bei et al, NIM A, 622(3) (2010) 518
- [10] Y. Cai, NIM A 645,168 (2011); Y. Cai et al, PRST-AB 14, 054002 (2012)

### 2.2.5 Betatron Motion with Coupling of Two Degrees of Freedom

*V. Lebedev, A. Burov, FNAL*

Betatron coupling was initially considered as an undesired effect, and efforts were made to suppress it. However, over recent three decades the betatron coupling has become an intrinsic part of several proposed and built accelerators [1]-[4]. (See also Sec.2.1.3.) Presently, two different basic representations are mostly used. The first was proposed by Edwards and Teng [5, 6] and the second by Mais and Ripken [7, 8]. The Mais-Ripken representation was later extended by Lebedev and Bogacz [9], where the equivalence of the two basic representations was established.

#### Motion equations and symplecticity condition

For a linear Hamiltonian system with two degrees of freedom, equations of motion can be written in the following matrix form:

$$\frac{dx}{ds} = \mathbf{U} \mathbf{H} \mathbf{x}, \quad \mathbf{U} = \begin{bmatrix} J & 0 \\ 0 & J \end{bmatrix}, \quad \mathbf{J} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \quad (1)$$

Here  $s = \beta ct$  is the conventional distance used as time variable,  $\mathbf{x} = (x, \theta_x, y, \theta_y)^t$  is a vector of canonical variables,  $\mathbf{H}$  is a Hessian matrix of

the Hamiltonian,  $\mathbf{H}_{ik} = \partial^2 H / \partial x_i \partial x_k$ . In the absence of longitudinal magnetic field,  $\theta_x = x'$  and  $\theta_y = y'$  are the horizontal and vertical angles; otherwise  $\theta_x = x' - Ry/2$ ,  $\theta_y = y' + Rx/2$ , where  $R = eB_s/Pc$ ,  $B_s$  is the longitudinal magnetic field and  $P$  is the total momentum. For a flat horizontal orbit,

$$\mathbf{H} = \begin{bmatrix} K^2 + k + \frac{R^2}{4} & 0 & N & -\frac{R}{2} \\ 0 & 1 & \frac{R}{2} & 0 \\ N & \frac{R}{2} - k + \frac{R^2}{4} & 0 & 0 \\ -R/2 & 0 & 0 & 1 \end{bmatrix} \quad (2)$$

where  $K = eB_y/Pc$ ,  $k = eG/Pc$ ,  $N = eG_s/Pc$ ,  $B_y$  is the vertical component of magnetic field;  $G$  and  $G_s$  are the normal and skew (tilted by  $+45^\circ$  around the  $s$  axis) components of the magnetic field gradient. Note that a quad tilted by an arbitrary angle can be represented by a linear combination of normal and skew quads. For any two solutions of Eq.(1),  $\mathbf{x}_1^t \mathbf{U} \mathbf{x}_2 = \text{const}$ . These constants are called the Lagrange invariants.

Let  $\mathbf{M}(0, s)$  be a transfer matrix:  $\mathbf{x}(s) = \mathbf{M}(0, s)\mathbf{x}(0)$ . Conservation of the Lagrange invariant yields that  $\mathbf{M}$  is symplectic:

$$\mathbf{M}^t \mathbf{U} \mathbf{M} = \mathbf{U} \Leftrightarrow \mathbf{M} \mathbf{U} \mathbf{M}^t = \mathbf{U} \quad (3)$$

This leaves only 10 independent parameters for 16 matrix elements. In a ring, its revolution matrix  $\mathbf{M}$  has four eigenvalues  $\lambda_i$  and eigenvectors  $\mathbf{v}_i$ ,

$$\mathbf{M} \mathbf{v}_i = \lambda_i \mathbf{v}_i, \quad i = 1, \dots, 4 \quad (4)$$

A stable betatron motion requires all eigenvalues to be confined to a unit circle, none of them equal to  $\pm 1$ ; thus, the eigenvalues are split into two complex conjugated pairs:  $\lambda_3 = \lambda_1^*$ ,  $\lambda_4 = \lambda_2^*$ , and similarly the eigenvectors. For any two normalized eigenvectors, condition (3) yields their symplectic orthonormality,

$$\begin{aligned} \mathbf{v}_l^\dagger \mathbf{U} \mathbf{v}_l &= -2i, \quad l = 1, 2 \\ \mathbf{v}_k^\dagger \mathbf{U} \mathbf{v}_m &= 0, \quad k \neq m = 1, \dots, 4 \end{aligned} \quad (5)$$

where  $\dagger$  stands for the Hermit conjugation. The choice of Eq.(5) normalization constants makes the matrix

$$\mathbf{V} = [\text{Rev}_1, -\text{Imv}_1, \text{Rev}_2, -\text{Imv}_2] \quad (6)$$

symplectic and selects one (prime) eigen-vector from each pair.

#### Eigen-vectors and ellipsoid in 4-D phase space

Turn-by-turn particle positions and angles can be described as

$$\begin{aligned} \mathbf{x}_n &= \text{Re} \left( A_1 e^{-i(\psi_1+n\mu_1)} \mathbf{v}_1 + A_2 e^{-i(\psi_2+n\mu_2)} \mathbf{v}_2 \right) \\ &= \mathbf{V} \mathbf{S}^n \xi_A \end{aligned} \quad (7)$$

where four real parameters,  $A_1$ ,  $A_2$ ,  $\psi_1$  and  $\psi_2$  represent the betatron amplitudes and phases, and

$$\mathbf{S} = \begin{bmatrix} \cos \mu_1 & \sin \mu_1 & 0 & 0 \\ -\sin \mu_1 \cos \mu_1 & 0 & 0 & 0 \\ 0 & 0 & \cos \mu_2 & \sin \mu_2 \\ 0 & 0 & -\sin \mu_2 \cos \mu_2 & 0 \end{bmatrix}$$

$$\xi_A = \begin{bmatrix} A_1 \cos \psi_1 \\ -A_1 \sin \psi_1 \\ A_2 \cos \psi_2 \\ -A_2 \sin \psi_2 \end{bmatrix} \quad (8)$$

The 3-D surface inscribing an ensemble of particles, whose turn-by-turn coordinates are contained in a 4-D ellipsoid, can be determined by the bilinear form

$$\mathbf{x}^t \Xi \mathbf{x} = 1 \quad (9)$$

It is related to the mode emittances  $\varepsilon_1$  and  $\varepsilon_2$  by [9]

$$\Xi = \mathbf{U} \hat{\Xi} \mathbf{V}^t \mathbf{U}^t \quad (10)$$

where

$$\hat{\Xi} = \text{diag}(1/\varepsilon_1, 1/\varepsilon_1, 1/\varepsilon_2, 1/\varepsilon_2) \quad (11)$$

and the function  $\text{diag}(\dots)$  forms a diagonal  $4 \times 4$  matrix. The inversion of Eq.(10) is

$$\hat{\Xi} = \mathbf{V}^t \Xi \mathbf{V} \quad (12)$$

i.e. a symplectic transform  $\mathbf{V}$  reduces matrix  $\Xi$  to its diagonal form. The beam emittance is a product of the semi-axes of the ellipsoid described by  $\hat{\Xi}$  so that

$$\varepsilon_{4D} = \left( \hat{\Xi}_{11} \hat{\Xi}_{22} \hat{\Xi}_{33} \hat{\Xi}_{44} \right)^{-1/2} = \left( \det \hat{\Xi} \right)^{-1/2}$$

$$= (\det \Xi)^{-1/2} = \varepsilon_1 \varepsilon_2 \quad (13)$$

where we omitted the factor  $\pi^2/2$  correcting for the actual 4-D volume of the ellipsoid. The quadratic form  $\Xi$  determines the emittances and eigenvectors,

$$(\Xi - i\varepsilon_l^{-1} \mathbf{U}) \mathbf{v}_l = 0; \quad \det(\Xi - i\varepsilon_l^{-1} \mathbf{U}) = 0 \quad (14)$$

For a Gaussian distribution

$$f(\mathbf{x}) = (4\pi^2 \varepsilon_1 \varepsilon_2)^{-1} \exp(-\mathbf{x}^t \Xi \mathbf{x}/2) \quad (15)$$

and the second-order moments are

$$\Sigma_{ij} \equiv \overline{x_i x_j} = \int x_i x_j f(\mathbf{x}) dx^4$$

$$= (\mathbf{V} \hat{\Xi}^{-1} \mathbf{V}^t)_{ij} = (\Xi^{-1})_{ij} \quad (16)$$

Consequently, a symplectic transform  $\mathbf{V}\mathbf{U}$  reduces matrix  $\Sigma$  to its diagonal form. Similarly to

Eq.(14), the emittances and the eigenvectors can be found from the matrix of moments  $\Sigma$ ,

$$\det(\Sigma \mathbf{U} + i\varepsilon_l \mathbf{I}) = 0; \quad (\Sigma \mathbf{U} + i\varepsilon_l \mathbf{I}) \hat{\mathbf{v}}_l = 0 \quad (17)$$

The mode emittances  $\varepsilon_1$  and  $\varepsilon_2$  are the motion invariants, i.e. they cannot be changed in the course of linear Hamiltonian motion. In particular, if initial and final states are uncoupled, the only possible emittance exchange is an emittance permutation. This statement is correct for three degrees of freedom as well.

For circular machines, the eigen-vectors  $\mathbf{v}_i$  are determined by the revolution matrix, Eq.(4). For transfer lines, they are determined by the phase space ellipsoid of the incoming beam, as given by Eqs.(14) or (17). The downstream eigenvectors and phase advances can be found with Eq.(7).

**Mais-Ripken parameterization** Symplecticity leaves only 8 independent real parameters for the eigenvectors of stable betatron motion. An extension [9] of Mais-Ripken (MR) parameterization [7] suggests the form

$$\mathbf{v}_1 = \left( \sqrt{\beta_{1x}}, -\frac{i(1-u)+\alpha_{1x}}{\sqrt{\beta_{1x}}}, \sqrt{\beta_{1y}} e^{i\nu_1}, -\frac{iu+\alpha_{1y}}{\sqrt{\beta_{1y}}} e^{i\nu_1} \right)^t$$

$$\mathbf{v}_2 = \left( \sqrt{\beta_{2x}} e^{i\nu_2}, -\frac{iu+\alpha_{2x}}{\sqrt{\beta_{2x}}} e^{i\nu_2}, \sqrt{\beta_{2y}}, -\frac{i(1-u)+\alpha_{2y}}{\sqrt{\beta_{2y}}} \right)^t$$

Here  $\beta_{ix,y}$ ,  $\alpha_{ix,y}$  are generalized Courant-Snyder functions; three other real functions,  $u$  and  $\nu_{1,2}$ , can be expressed in terms of the Courant-Snyder functions. The transfer matrix  $\mathbf{M}(s_1, s_2)$  between two points can be expressed through the eigenvectors of these points,  $\mathbf{V}(s_1)$  and  $\mathbf{V}(s_2)$ :

$$\mathbf{M}(s_1, s_2) = -\mathbf{V}(s_2) \mathbf{S} \mathbf{U} \mathbf{V}(s_1)^t \mathbf{U}$$

$$= \mathbf{V}(s_2) \mathbf{S} \mathbf{V}(s_1)^{-1} \quad (18)$$

where the matrix  $\mathbf{S}$  is given by Eq.(8). Since all matrices in Eq.(18) are symplectic, this equation yields a symplectic transfer matrix for any incoming and outgoing eigenvectors. Consequently, the transformation  $\mathbf{M}(s_1, s_2)$  can always be implemented with conventional linear optics elements.

**Edwards-Teng parameterization** Edwards-Teng (ET) parameterization [6] is based on a canonical transform which reduces a  $4 \times 4$  revolution matrix  $\mathbf{M}$  to its block-diagonal normal modes form,

$$\mathbf{M} = \mathbf{R} \hat{\mathbf{M}} \mathbf{R}^{-1}, \quad \hat{\mathbf{M}} = \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{B} \end{bmatrix} \quad (19)$$

Matrices  $\mathbf{A}$  and  $\mathbf{B}$  are parameterized similar to the uncoupled motion. Thus, matrix  $\hat{\mathbf{M}}$ , describing particle motion in new coordinates, is parameterized by six real parameters. A symplectic matrix  $\mathbf{R}$  is parameterized as

$$\mathbf{R} = \begin{bmatrix} \mathbf{I} \cos \varphi & \mathbf{D}^{-1} \sin \varphi \\ -\mathbf{D} \sin \varphi & \mathbf{I} \cos \varphi \end{bmatrix} \quad (20)$$

where  $\mathbf{I}$  is the unit  $2 \times 2$  matrix, and  $\mathbf{D}$  is a  $2 \times 2$  symplectic matrix. Consequently, ten parameters fully describe the transfer matrix  $\mathbf{M}$ . They are the six parameters of the decoupled motion,  $\beta_{1,2}$ ,  $\alpha_{1,2}$ ,  $\mu_{1,2}$ , Teng's phase  $\phi$ , and three parameters of matrix  $\mathbf{D}$ . Their relation to the MR parameters can be found in [9], in particular  $\sin \varphi = \pm \sqrt{u}$ ,  $\beta_{1,2} = \beta_{1x,2y}/(1-u)$ ,  $\alpha_{1,2} = \alpha_{1x,2y}/(1-u)$ . For ET, the phase advances of the betatron motion are defined as

$$\hat{\mathbf{v}}_i(s) e^{-i\mu_i(s)} = \hat{\mathbf{M}}(0, s) \hat{\mathbf{v}}_i(0) \quad (21)$$

where  $\hat{\mathbf{v}}_i(s)$  are the decoupled eigenvectors in the normal mode coordinates. Comparison of Eq.(19) with Eq.(18) shows that ET and MR present the same eigen-modes with different sets of parameters. Therefore their phase advances are identical.

Usually optical codes support an analysis of coupled optics using either MR or ET formalism: MAD8 [13] supports MR and ET; MAD-X [14] supports ET, and OptiM [15] supports MR.

**Round beams and beam adapters** Some applications require round (axially-symmetric) beams. These beams are conveniently described in a basis of circular modes, which are given by Eq.(18) with  $\beta_{lx} = \beta_{ly} = \beta$ ,  $\alpha_{lx} = \alpha_{ly} = \alpha$ ,  $u = 1/2$ , and  $\nu_{1,2} = \pm \pi/2$ . In a matched solenoid  $\beta = 1/R$  and one of modes is a Larmor motion with center at the solenoid axis, and another one is a pure offset,  $x, y = \text{constant}$ . A product of mode emittances is equal to the total 4-D emittance (see Eq.(13)), and their difference is the beam angular momentum  $\langle x\theta_y - y\theta_x \rangle$ . If a linear transformation  $\mathbf{M}$  preserves axial symmetry, it also preserves an absolute value of the canonical angular momentum of every particle. The general form of this transformation is

$$\mathbf{M} = \mathbf{R}(\theta) \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \pm \mathbf{A} \end{bmatrix} \quad (22)$$

where  $\mathbf{R}(\theta)$  describes a rotation with an arbitrary angle in both  $x, y$  and  $\theta_x, \theta_y$  subspaces, and  $\mathbf{A}$  is an arbitrary  $2 \times 2$  symplectic matrix. A transformation cannot change an absolute value

of the particle canonical angular momentum without breaking the rotational symmetry of the entire ensemble [11]. It was found by Derbenev [1] that a conventional planar beam state, with  $\beta_{2x} = \beta_{1y} = \alpha_{2x} = \alpha_{1y} = u = 0$ , can be transformed into a round beam and back by proper optical schemes, called him as beam adapters. (See also Sec.2.1.3) To illustrate the main idea, let eigenstates of the incoming beam be represented by two circular modes described above with  $\alpha_{lx,y} = 0$ , and  $\beta_{lx,y} = \beta$ . Let this beam undergo an uncoupled transformation  $\mathbf{T}$ , which horizontal part is the identity transformation  $\mathbf{I}$ , and vertical part gives  $90^\circ$  phase advance,

$$\mathbf{T} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{N} \end{bmatrix}; \quad \mathbf{N} = \begin{bmatrix} 0 & 2\beta \\ -(2\beta)^{-1} & 0 \end{bmatrix}$$

It is straightforward to check that this transformation converts two circular modes into two planar modes tilted by  $45^\circ$ . Since the initial beam state is axially symmetric, the transfer line tilt by  $45^\circ$  rotates the modes making them conventional planar modes. The simplest adapter implementation requires a system of three quadrupoles [10].

**Perturbation theory** Let the revolution matrix  $\mathbf{M}$  be perturbed:  $\mathbf{M} = (\mathbf{I} + \mathbf{P})\mathbf{M}_0$ , where the perturbation  $\mathbf{P}$  is small, but not necessarily symplectic. In the first order of perturbation theory, the complex tune shifts are [16]

$$\Delta\mu_l/(2\pi) = -(4\pi)^{-1} \mathbf{v}_l^\dagger \mathbf{U} \mathbf{P} \mathbf{v}_l, \quad l = 1, 2 \quad (23)$$

In particular, this formula allows calculation of the incoherent beam-beam and space charge tune shifts for arbitrary-coupled optics [17]. It also yields that a sum of the two growth rates is independent on the eigenvectors,

$$\text{Im}(\Delta\mu_1 + \Delta\mu_2) = \text{Tr}(\mathbf{P})/2 \quad (24)$$

This statement is sometimes called as a rate-sum theorem.

**Coherent motion** There is a general rule reducing a problem of coherent motion for arbitrary coupled optics to uncoupled one. If the growth rates are smaller than the tune difference the two problems are analytically identical for any beam distribution function, rf configuration and impedances  $Z_{x,y}$ , with the following rule of correspondence [18]:

$$\beta_{x,y} Z_{x,y} \rightarrow \beta_{lx} Z_x + \beta_{ly} Z_y; \quad l = 1, 2 \quad (25)$$

Coupling adjustments also affect the chromaticities of betatron modes and their space charge tune shifts; all these factors make beam stability non-trivially dependent on coupling [19].

**Acceleration** Hereinabove, no acceleration is assumed; however, acceleration is not actually an obstacle to the described formalism. Indeed, let  $P_0$  and  $P$  be the initial and final momenta. Then the transfer matrices are symplectic in the vector space  $\tilde{\mathbf{x}} = \sqrt{P/P_0} \mathbf{x}$ , making the above formalism applicable to the case of beam acceleration.

**Applications** Applications of coupled optics formalism and the planar-circular adapters include relativistic electron cooling, round beams for circular colliders, flat electron beams for linear colliders, low-energy hadron storage rings, and ionization cooling of muons [11, 12].

**FNAL Electron cooling** Because of very high electron beam energy, the beam transport in Fermilab electron cooler uses conventional optics elements instead of normally used longitudinal magnetic field [2]. To keep the electron beam focused in the cooling section and minimize its effective temperature, both the cathode and the cooling section are immersed in the longitudinal magnetic field, so that the magnetic fluxes through the beam at the cathode and cooler are identical. It requires the beam transport from cathode to solenoid to be rotationally invariant, Eq.(22). The cooling section is 20 m long, and the total length of beam transport (from cathode to collector) is ~100 m. In addition to solenoids, the beam transport includes 10 dipoles, which do not provide an axially-symmetric transformation. However, with appropriate strengths and locations of focusing solenoids, the transport map from the cathode to the cooling section and back to the collector is rotationally invariant. Matching electron and antiproton beam radii in the cooler imposes an additional requirement to the cathode-cooler electron supply line [11].

**Neutrino factories and muon colliders** require 3-D ionization cooling of muons. It cannot be achieved without coupling between degrees of freedom. To maximize the cooling range and efficiency, most schemes use solenoidal focusing which introduces  $x$ - $y$  coupling at the fundamental level. To couple the transverse and longitudinal degrees of freedom, one needs to excite the dispersion, i.e. to make a curvilinear beam orbit [20, 21]. The resonance dispersion excitation

proposed in [22] reduces the tune chromaticity in vicinity of integer resonance resulting in an increase of the cooling channel momentum acceptance.

**Round beams for circular colliders** were proposed [23, 24] and successfully implemented [25, 26] to mitigate the beam-beam effects. If the revolution map in the interaction point (IP) is rotationally invariant, than the eigen-vectors are rotationally invariant too ( $\beta_{1x} = \beta_{2x} = \beta_{1y} = \beta_{2y}$ ,  $\alpha_{1x} = \alpha_{2x} = \alpha_{1y} = \alpha_{2y}$ ,  $u = 1/2$ ); and the betatron tunes are equal to  $\mu_{1,2} = \mu \pm \theta$ , where  $\mu$  is the betatron phase advance for matrix  $\mathbf{A}$  in Eq.(22). In this case, the angular momentum is preserved in the course of beam transport. For  $\theta = 0$  the fractional parts of both tunes  $\nu_{1,2} = \mu_{1,2}/(2\pi)$  are equal, allowing to avoid dangerous beam-beam resonances. Note that for  $\theta = 0$  the map rotational invariance does not automatically results in the beam being round. The degeneracy of eigenmodes yields that they can be chosen circular or planar with arbitrary rotation angle. In the first case the beam is round for any ratio of the mode emittances, while in the second case only for equal mode emittances. If the beam is not round, the beam-beam collisions do not preserve rotational symmetry, and the axially-symmetric optics does not bring significant benefits for beam-beam effects suppression. In reality, the block-diagonal  $x$  and  $y$  components  $\mathbf{A}$  in the map (22) cannot be exactly identical; differences in the corresponding Twiss parameters  $\beta_{x,y}$  and phase advances  $\mu_{x,y}$  remove degeneracy and make modes planar for  $\theta = 0$ . To provide mode circularity, the solenoidal coupling angle  $\theta$  has to exceed uncoupled mismatch,

$$|\theta| \gg |\mu_x - \mu_y|, \quad |1 - \beta_x/\beta_y|$$

To get round beams in IP of an  $e^+e^-$  collider, R. Talman suggested the Möbius revolution map [27]

$$\mathbf{M} = \begin{bmatrix} \mathbf{0} & -\mathbf{Y} \\ \mathbf{X} & \mathbf{0} \end{bmatrix} \quad (26)$$

where  $\mathbf{X}$  and  $\mathbf{Y}$  are  $2 \times 2$  matrices with arbitrary phase advances  $\mu_x$  and  $\mu_y$ . Experimental investigation of that lattice at CESR showed that it is not as successful as the axially-symmetric optics [25]. Indeed, for  $\mu_x \neq \mu_y$  the map (26) does not belong to the rotation-invariant group (22) and does not preserve the angular momentum. Although the two emittances are equal for that lattice, advantages of the rotation-invariant optics do not apply.

Instead, the axially-symmetric Möbius map,  $\mu_x = \mu_y$ ,  $\beta_x = \beta_y$ ,  $\alpha_x = \alpha_y$ , preserves angular momentum, so from that point of view it should be as good as uncoupled axially-symmetric optics. The axially-symmetrical Möbius corresponds to  $\theta = \pi/2$ ; thus, the tunes are different by  $\frac{1}{2}$ . Although the beam-beam resonance suppression may be as good as for  $\theta = 0$ , it still requires an experimental verification. The Möbius optics also requires a redesign of a chromaticity correction.

## References

- [1] Ya. Derbenev, U. Michigan, UM HE 93-20 (1993); UM HE 98-04 (1998)
- [2] S. Nagaitsev et al, PRL 96, 044801 (2006)
- [3] R. Talman, PRL 74, 1590 (1995)
- [4] R.C. Fernow et al, Proc. AIP 372, 146 (1995)
- [5] L.C. Teng, Fermilab, FN-229 0100 (1971)
- [6] D.A. Edwards, L.C. Teng, IEEE Trans. Nucl. Sci. 20, 3, p.885 (1973)
- [7] I. Borchardt et al, DESY 87-161
- [8] F. Willeke, G. Ripken, AIP Conf. Proc. 184 (1989)
- [9] V. Lebedev, S. Bogacz, e-print JLAB-ACC-99-19 (1999), JINST, v.5 (2010)
- [10] A. Burov, V. Danilov, Fermilab-TM-2043 (1998)
- [11] A. Burov et al, PRST-AB 3, 094002 (2000)
- [12] A. Burov, S. Nagaitsev, Ya. Derbenev, PRE 66, 016503 (2002)
- [13] <http://mad.web.cern.ch/mad/mad8web/mad8.html>
- [14] [http://mad.web.cern.ch/mad/madx\\_manual.pdf](http://mad.web.cern.ch/mad/madx_manual.pdf)
- [15] <http://www-bdnew.fnal.gov/pbar/organizationalchart/lebedev/OptiM/optim.htm>
- [16] A. Burov, PRST-AB 9, 120101 (2006)
- [17] A. Burov, Ya. Derbenev, Fermilab-PUB-09-392-AD (2010)
- [18] A. Burov, V. Lebedev, PRST-AB 10, 044402 (2007)
- [19] R. Cappi, E. Métral, D. Möhl, CERN/PS 2001-010 (AE) (2001)
- [20] Ya. Derbenev, R. Johnson, PRST-AB 8, 041002 (2005)
- [21] R.C. Fernow, R.B. Palmer, PRST-AB 10, 064001 (2007)
- [22] Y. Alexahin, AIP Conf. Proc. 1222, 313 (2010)
- [23] S. Krishnagopal, R. Siemann, PAC'89, p.836
- [24] V. Danilov et al, EPAC'96, p.1149
- [25] S. Henderson et al, PAC'99, p.410
- [26] D. Berkava et al, ICFA Beam Dyn. Newslett. 48, 235 (2009)
- [27] R. Talman, PRL 24, 1590 (1995)

## 2.3 NONLINEAR DYNAMICS

### 2.3.1 Hamiltonian

*K. Symon, U. Wisconsin*

#### 2.3.1.1 General case

The Hamiltonian for a particle of charge  $e$ , mass  $m$ , moving in an electromagnetic field described by electromagnetic potentials  $\vec{A}_{\text{em}}$ ,  $\phi$  is [1, 2, 3, 4, 6, 7]

$$H_t(x, p_x, y, p_y, s, p_s; t) = e\phi + T \quad (1)$$

where

$$\begin{aligned} T = c & \left[ m^2 c^2 + (p_x - eA_x)^2 + (p_y - eA_y)^2 \right. \\ & \left. + \left( \frac{p_s - eA_s}{1 + \rho^{-1}x} \right)^2 \right]^{1/2} = \gamma mc^2 \end{aligned} \quad (2)$$

is the kinetic energy. The coordinates are in Fig.1. The canonical vector potentials are

$$\begin{aligned} A_x &= \hat{x} \cdot \vec{A}_{\text{em}} \\ A_y &= \hat{y} \cdot \vec{A}_{\text{em}} \\ A_s &= \left( 1 + \frac{x}{\rho} \right) \hat{s} \cdot \vec{A}_{\text{em}} \end{aligned} \quad (3)$$

For a circular accelerator, all quantities are periodic functions of  $s$ .

With  $s$  as the independent variable the Hamiltonian is

$$H_s(x, p_x, y, p_y, t, -E; s) = -eA_s - \left( 1 + \frac{x}{\rho} \right) P_s \quad (4)$$

where  $-E$  is the momentum conjugate to the coordinate  $t$  and

$$\begin{aligned} P_s = & \left[ \left( \frac{E - e\phi}{c} \right)^2 - (p_x - eA_x)^2 \right. \\ & \left. - (p_y - eA_y)^2 - m^2 c^2 \right]^{1/2} = \beta_s \gamma mc \end{aligned} \quad (5)$$

is the kinetic momentum in the  $s$  direction.

The canonical potentials for magnetic lattice elements, expanded to 4th order in  $x, y$ , are:

$$\text{straight section: } \vec{A} = 0, \quad (6)$$

$$\begin{aligned} \text{bend: } \vec{A} &= \hat{s} \left[ -Bx + \left( \frac{B_1 \sin \alpha}{2} - \frac{B}{2\rho} \right) x^2 \right. \\ &\quad - \frac{B_1 \sin \alpha}{2} y^2 + \left( \frac{B_1 \sin \alpha}{3\rho} - \frac{B_2 \sin^2 \alpha}{6} \right) x^3 \\ &\quad + \left( \frac{B_2 \sin^2 \alpha}{2} - \frac{B_1 \sin \alpha}{2\rho} \right) xy^2 \\ &\quad + \left( \frac{B_3 \sin^3 \alpha}{24} - \frac{B_2 \sin^2 \alpha}{8\rho} \right) x^4 \\ &\quad + \left( \frac{B_2 \sin^2 \alpha}{2\rho} - \frac{B_3 \sin^3 \alpha}{4} \right) x^2 y^2 + \frac{B_3 \sin \alpha}{24} y^4 \Big] \\ &\quad + \hat{x} \left[ -\frac{B_1 \cos \alpha}{2} y^2 + \frac{B_2 \sin \alpha \cos \alpha}{2} xy^2 \right. \\ &\quad \left. - \frac{B_3 \sin^2 \alpha \cos \alpha}{4} x^2 y^2 + \frac{B_3 \cos \alpha}{24} y^4 \right] \end{aligned} \quad (7)$$

$$\begin{aligned} \text{gradient bend: } \vec{A} &= \hat{s} \left[ -Bx - \left( \frac{B}{2\rho} + \frac{G}{2} \right) x^2 \right. \\ &\quad + \frac{G}{2} y^2 - \frac{G}{3\rho} x^3 + \frac{G}{2\rho} xy^2 \\ &\quad + \left( \frac{B'^2}{24B\rho} + \frac{B''}{12\rho} - \frac{G''}{24} + \frac{G}{24\rho^2} \right) y^4 \Big] \\ &\quad + \hat{x} \left[ -\frac{B'}{2} y^2 + \left( \frac{B'}{2\rho} - \frac{G'}{2} \right) xy^2 \right. \\ &\quad \left. - \left( \frac{B'}{2\rho^2} - \frac{G'}{2\rho} \right) x^2 y^2 + \left( \frac{B'''}{24} + \frac{G'}{24\rho} \right) y^4 \right] \end{aligned} \quad (8)$$

$$\begin{aligned} \text{quad: } \vec{A} &= \hat{s} \left[ \frac{Q}{2} (y^2 - x^2) \right. \\ &\quad \left. + \frac{Q''}{48} (x^4 - 6x^2 y^2 - y^4) \right] - \hat{x} \frac{Q'}{2} xy^2 \end{aligned} \quad (9)$$

$$\begin{aligned} \text{skew quad: } \vec{A} &= \hat{s} \left[ Q_s xy - \frac{Q''_s}{6} x^3 y \right] \\ &\quad + \hat{x} Q'_s \left( \frac{y^3}{6} - \frac{x^2 y}{2} \right) \end{aligned} \quad (10)$$

$$\begin{aligned} \text{sextupole: } \vec{A} &= \hat{s} S \left( \frac{1}{2} xy^2 - \frac{1}{6} x^3 \right) \\ &\quad + \hat{x} S' \left( -\frac{1}{4} x^2 y^2 + \frac{1}{24} y^4 \right) \end{aligned} \quad (11)$$

$$\text{octupole: } \vec{A} = \hat{s} O \left( \frac{1}{4} x^4 - \frac{3}{2} x^2 y^2 + \frac{1}{4} y^4 \right) \quad (12)$$

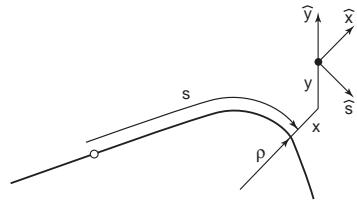


Figure 1: Coordinates Relative to Reference Orbit:  $x$  is distance measured horizontally outward,  $y$  measured vertically,  $s$  measured along the reference orbit from some reference point, and  $\rho$  is the local radius of curvature of the reference orbit, which lies in the  $x$ - $s$  plane.

defined near the reference orbit by:

$$\text{bend: } \vec{B} = \hat{y} B(s) \quad (13)$$

$$\text{grad. bend: } \vec{B} = \hat{y} [B(s) + G(s)x] \quad (14)$$

$$\text{quad: } \vec{B} = \hat{y} Q(s)x + \hat{x} Q(s)y \quad (15)$$

$$\text{skew quad: } \vec{B} = \hat{x} Q_s(s)x - \hat{y} Q_s(s)y \quad (16)$$

$$\begin{aligned} \text{sextupole: } \vec{B} &= \hat{y} \frac{S(s)}{2} (x^2 - y^2) \\ &\quad + \hat{x} S(s)xy \end{aligned} \quad (17)$$

$$\begin{aligned} \text{octupole: } \vec{B} &= \hat{x} O(s)(y^3 - 3x^2 y) \\ &\quad - \hat{y} O(s)(x^3 - 3y^2 x) \end{aligned} \quad (18)$$

In Eq.(7),  $\alpha(s)$  is the angle between the local tangent to the reference orbit and the perpendicular to the magnet edge. The parameters in Eq.(7) are given by the field and its gradients on the reference orbit,

$$B_1 = \frac{B'}{\cos \alpha} \quad (19)$$

$$B_2 = \frac{B''}{\cos^2 \alpha} - \frac{B' \sin \alpha}{\rho \cos^3 \alpha} \quad (20)$$

$$\begin{aligned} B_3 &= \frac{B'''}{\cos^3 \alpha} - \frac{3B'' \sin \alpha}{\rho \cos^4 \alpha} \\ &\quad + \frac{B'(1+2 \sin^2 \alpha)}{\rho^2 \cos^5 \alpha} - \frac{B'^2 \sin \alpha}{B \rho \cos^4 \alpha} \end{aligned} \quad (21)$$

Eq.(8) is for a bend with a gradient. Primes denote derivatives with respect to  $s$  and appear in edge corrections.

Edge corrections depend on the edge design. Eq.(7) assumes a straight edge (no field parallel to the edge and no dependence of the fields on the coordinate parallel to the edge). It is accurate if edge fields do not extend past the magnet center. For a thin edge perpendicular to the reference orbit,  $\alpha = 0$ . Eq.(8) assumes the magnet contours run perpendicular to the reference orbit at each point, so that in the median plane the

where  $B(s), G(s), Q(s), S(s), O(s)$  are the vertical field and its gradients on the reference orbit,

field gradient depends only on  $s$ . The formulae for multipoles assume the edge fields preserve the multipole symmetry about the reference orbit. For other edge designs, these formulae should give approximate values of the coefficients, particularly if the edges are thin, but they do not include any terms which violate the assumed symmetries.

The potential for an accelerating gap may be written, to second order,

$$\vec{A} = - \int^t V(t) dt \left\{ \hat{s} \left[ \Delta(s) - \frac{\Delta''}{4}(x^2 + y^2) \right] - \frac{\Delta'}{2}(\hat{x}\hat{x} + \hat{y}\hat{y}) \right\} + \hat{s} \frac{\Delta c^{-2}\dot{V}}{4}(x^2 + y^2) \quad (22)$$

where  $V(t)$  is gap voltage;  $\Delta(s)$  describes the field shape with  $\int_{\text{gap}} ds \Delta(s) = 1$ . Usually, one approximates  $\Delta(s) = \delta(s - s_{\text{gap}})$ . The terms in  $x, y$  are edge field corrections appropriate to a circular gap perpendicular to the reference orbit; they are usually neglected.

The potentials for space charge fields are

$$\phi = \frac{(1-f)\lambda}{4\pi\epsilon_0} g(x, y) \quad (23)$$

$$\vec{A} = \hat{s} \frac{\beta\lambda}{4\pi\epsilon_0 c} (1 + \frac{x}{\rho}) g(x, y) \quad (24)$$

where  $\lambda$  is charge line density,  $f$  is the neutralization fraction. Time rates of change are assumed slow relative to light transit time across the vacuum chamber. For a uniform cylindrical charge of radius  $a$ , centered in a circular perfect conductor of radius  $b$ , the function  $g(x, y)$  is ( $r^2 = x^2 + y^2$ )

$$g(x, y) = \begin{cases} 1 + 2 \ln \frac{b}{a} - \frac{r^2}{a^2} & \text{if } r \leq a \\ \ln \frac{b^2}{r^2} & \text{if } a \leq r \leq b \end{cases} \quad (25)$$

$a$  and  $b$  may be functions of  $s$  and  $t$ . Some other cases are found in [2, 3, 5].

### 2.3.1.2 Transverse motion

From Eqs.(4)-(12), we obtain the Hamiltonian for transverse motion:

$$H_{sT}(x, p_x, y, p_y; s) = -\frac{eA_s}{p} - (1 + \frac{x}{\rho}) \left[ 1 - \left( p_x - \frac{eA_x}{p} \right)^2 - p_y^2 \right]^{\frac{1}{2}} \quad (26)$$

where the kinetic momentum is

$$p = \left( \frac{E^2 - m^2 c^4}{c^2} \right)^{\frac{1}{2}} \quad (27)$$

and we have rescaled  $p_x, p_y, H_{sT}$  to be  $1/p$  (not  $1/p_0$ ) times their values in Eq.(4). The electric potential  $\phi$  has been dropped. Space charge can be included by adding

$$A_s = - \left( 1 + \frac{x}{\rho} \right) \left( \frac{1}{\gamma^2} - f \right) \frac{\lambda g(x, y)}{4\pi\epsilon_0 c \beta} \quad (28)$$

which is not the canonical potential, but gives the space charge field including the  $\phi$  term under the conditions of Eqs.(23)-(25).

The expansion of Eq.(26) to fourth order in  $x, y, p_x, p_y$  is

$$\begin{aligned} H_{sT} = & \frac{1}{2} p_x^2 + \frac{1}{2} p_y^2 - \frac{x}{\rho} \left( 1 - \frac{p_0}{p} \right) \\ & - \sum_{k+\ell=2}^4 \frac{e A_{sk\ell}}{p} x^k y^\ell + \frac{e B'}{2p} y^2 p_x \\ & + \frac{x(p_x^2 + p_y^2)}{2\rho} + \frac{e^2 B'^2}{8p^2} y^4 \\ & + \left( \frac{e B'}{2p\rho} - \frac{e A_{x12}}{p} \right) x y^2 p_x - \frac{e A_{x21}}{p} x^2 y p_x \\ & - \frac{e A_{x03}}{p} y^3 p_x + \frac{1}{8} (p_x^2 + p_y^2)^2 \\ & + C_{sc}(x^2 + y^2) + \frac{C_{sc}}{\rho} (x^3 + x y^2) \end{aligned} \quad (29)$$

where  $p_0 = eB\rho$  is the momentum of the reference particle,  $A_{sk\ell}$  and  $A_{xk\ell}$  are the coefficients multiplying  $x^k y^\ell$  in the appropriate terms of Eqs.(7)-(12), and

$$C_{sc} = \left( \frac{1}{\gamma^2} - f \right) \frac{e\lambda}{4\pi\epsilon_0 a^2 \beta p} \quad (30)$$

Only linear space charge terms have been included.

#### 2.3.1.3 Longitudinal motion

The Hamiltonian, to second order in  $w = E - E_s$  [2] is

$$H_{sL}(\varphi, w; s) = \frac{K}{2} w^2 + e \hat{V}(\cos \varphi + \varphi \sin \varphi_s) \Delta(s) \quad (31)$$

Subscript  $s$  refers to the synchronous energy (e.g.  $E_s, p_s, \phi_s, \gamma_s$ ),  $K = \frac{\omega_{rf} \eta}{c \beta^3 E_s}$ ,  $\hat{V}$  is gap voltage amplitude, and

$$\Delta(s) = \sum_j \delta(s - s_j) \quad (32)$$

For a linac,  $\eta = -\gamma^{-2}$ . For a circular accelerator,  $\eta = \gamma_t^{-2} - \gamma^{-2}$ ,  $s_j = s_{\text{gap}} + 2\pi R j$ .

The Hamiltonian (31) implies the difference equations

$$\varphi_{j+1} = \varphi_j + (s_{j+1} - s_j)K_j w_j \quad (33)$$

$$w_{j+1} = w_j + e\hat{V}_{j+1}(\sin \varphi_{j+1} - \sin \varphi_{s,j+1})$$

For a linac, if  $\hat{V}$  and the distance  $\Delta s$  between gaps is a smooth function of  $s$ , we may replace  $\Delta(s)$  by its average value,

$$\begin{aligned} H_{SL}(\varphi, w; s) &= \frac{1}{2}Kw^2 + \frac{e\hat{V}}{\Delta s}(\cos \varphi + \varphi \sin \varphi_s) \\ &\quad - \left( \frac{1}{\gamma^2} - f \right) \frac{e\omega_{rf}g_0}{4\pi\epsilon_0\beta c} \lambda(\varphi) \end{aligned} \quad (34)$$

$\lambda(\varphi)$  is line charge density at the point in the beam with phase  $\varphi$ , and  $g_0 = g(0, 0) = 1 + 2 \ln \frac{b}{a}$ .

For a circular accelerator, with time as the independent variable, the smoothed Hamiltonian is

$$\begin{aligned} H_{tL}(\varphi, w; t) &= \frac{1}{2}\beta c K w^2 + \frac{e\hat{V}\omega_0}{2\pi}(\cos \varphi + \varphi \sin \varphi_s) \\ &\quad - \left( \frac{1}{\gamma^2} - f \right) \frac{e\omega_{rf}g_0}{4\pi\epsilon_0} \lambda(\varphi) \end{aligned} \quad (35)$$

where  $\hat{V}$  is the maximum voltage gain per turn.

### 2.3.1.4 Synchrobetatron coupling

The complete Hamiltonian for synchrotron and horizontal betatron motion is [8, 1]

$$\begin{aligned} H_s(\underline{x}, \underline{p}_x, t, -E; s) \\ = -eA_s - p \left( 1 + \frac{\underline{x}}{\rho} \right) \left[ 1 - \frac{\underline{p}_x^2}{p^2} \right]^{1/2} \end{aligned} \quad (36)$$

where  $p$  is given by Eq.(27),  $\underline{x}$  is the horizontal distance measured from the reference orbit, and  $\underline{p}_x$  is the corresponding momentum.

Make a canonical transformation to synchrotron, betatron variables:

$$\begin{aligned} \phi &= - \int_0^t \omega_{rf}(t) dt + \frac{h(s - s_{gap})}{R} \\ &\quad + \frac{hD'}{R}x - \frac{hD}{Rp_s}p_x \end{aligned} \quad (37)$$

$$W = \frac{E - E_s}{\omega_{rf}} \quad (38)$$

$$x = \underline{x} - \frac{hD}{Rp_s}W \quad (39)$$

$$p_x = \underline{p}_x - \frac{hD'}{R}W \quad (40)$$

Subscript  $s$  denotes synchronous particle. From here on  $p_s = p_0(t = t_s)$ . Dispersion function  $D(s)$  satisfies  $D'' + kp_s D = 1/\rho$ , where

$p_s$  is treated as a constant and  $k(s) = \frac{1}{\rho^2} - \frac{Q}{p_0^2}$ . We assume gradients  $Q(s)$ , like bend fields  $B(s)$ , change slowly with time in proportion to  $p_0(t)$ .

The Hamiltonian for synchrobetatron coupling is

$$\begin{aligned} H_{SBC}(\phi, W, x, p_x; s) &= \frac{h^2}{2R^2p_s} \left( \frac{1}{\gamma_s^2} - \frac{D}{\rho} \right) W^2 \\ &\quad - \frac{e\hat{V}_s}{\omega_{rf}s} \Delta(s) \cos \left( \phi - \frac{hD'}{R}x + \frac{hD}{Rp_s}p_x \right) \\ &\quad - \frac{e\hat{V}_s}{2\pi R \omega_{rf}s} \left( \phi - \frac{hD'}{R}x + \frac{hD}{Rp_s}p_x \right) \sin \phi_s \\ &\quad + \frac{1}{2p_s} p_x^2 + \frac{kp_s}{2} x^2 \\ &\quad + H_{NLS} + H_{NLB} + H_{NLC} + H_{SVC} \end{aligned} \quad (41)$$

where  $h$  = harmonic number,  $\phi_s$  is defined by

$$\frac{dE_s}{ds} = - \frac{e\hat{V}_s}{2\pi R} \sin \phi_s \quad (42)$$

and the voltage across the accelerating gap is

$$V(t) = \hat{V}(t) \sin \int_0^t \omega_{rf}(t) dt \quad (43)$$

We have (i) assumed Eq.(32), (ii) assumed the bend edges  $\perp$  the reference orbit ( $\alpha = 0$ ), (iii) included quadrupoles, sextupoles, and straights, (iv) kept nonlinear terms up to third degree in  $W, x, p_x, \phi - \phi_s$ , and the time derivatives of the slowly varying quantities  $p_s, \hat{V}$ . The nonlinear synchrotron term is

$$\begin{aligned} H_{NLS} &= \left[ \left( \frac{h^3 D'^2}{2R^3 p_s^2} + \frac{h m^2 \omega_{rf}^2}{2R p_s^4} \right) \left( \frac{D}{\rho} - 1 \right) \right. \\ &\quad \left. + \frac{e S h^3 D^3}{6 R^3 p_s^3} \right] W^3 \end{aligned} \quad (44)$$

The cubic terms in the betatron variables are

$$H_{NLB} = \frac{e S}{6} x^3 + \frac{1}{2 p_s \rho} x p_x^2 \quad (45)$$

The nonlinear coupling terms are

$$\begin{aligned} H_{NLC} &= \frac{e S h D}{2 R p_s} x^2 W \\ &\quad + \left( \frac{e S h^2 D^2}{2 R^2 p_s^2} + \frac{\omega_{rf}^2 m^2}{2 p_s^3 \rho} + \frac{h^2 D'^2}{2 R^2 p_s} \right) x W^2 \\ &\quad + \frac{h D'}{R p_s \rho} x p_x W + \frac{h}{2 R p_s^2} \left( \frac{D}{\rho} - 1 \right) p_x^2 W \\ &\quad + \frac{h^2 D'}{R^2 p_s^2} \left( \frac{D}{\rho} - 1 \right) p_x W^2 \end{aligned} \quad (46)$$

The coupling terms linear in the time derivatives of slowly varying quantities are

$$\begin{aligned}
 H_{\text{SVC}} = & \frac{e\dot{\hat{V}}_s}{\omega_{\text{rfs}}^2} \Delta(s) \left\{ -\sin \underline{\phi} + (\underline{\phi} - \phi_s) \cos \underline{\phi} \right\} \\
 & - \frac{\dot{p}_s}{p_s \omega_{\text{rfs}}} \left\{ \frac{p_s}{\rho} x(\underline{\phi} - \phi_s) - \frac{h^2 D}{R^2 p_s} p_x W \right. \\
 & + \frac{h}{R} \left( \frac{D}{\rho} - \frac{1}{\gamma_s^2} \right) W(\underline{\phi} - \phi_s) - \frac{h^3 D D'}{2 R^3 p_s} W^2 \\
 & + \frac{e\dot{V}_s m^2 c^4}{\omega_{\text{rfs}} E_s^2} \left[ -\Delta(s) \sin \underline{\phi} \right. \\
 & \left. + \Delta(s)(\underline{\phi} - \phi_s) \cos \underline{\phi} + \frac{\sin \phi_s}{4\pi R} (\underline{\phi} - \phi_s)^2 \right] \right\} \quad (47)
 \end{aligned}$$

where

$$\underline{\phi} = \phi - \frac{h D'}{R} x + \frac{h D}{R p_s} p_x \quad (48)$$

Some handy formulae:

$$\frac{dt_s}{ds} = \frac{E_s}{p_s c^2} = \frac{1}{\beta_s c} \quad (49)$$

$$\omega_{\text{rfs}} = h \frac{\beta_s c}{R} = \frac{h p_s c^2}{R E_s} \quad (50)$$

$$\dot{\omega}_{\text{rfs}} = \frac{\omega_{\text{rfs}} m^2 c^4 p_s}{p_s E_s^2} \quad (51)$$

$$p = p_s + \frac{h}{R} W - \frac{\omega_{\text{rfs}}^2 m^2}{2 p_s^3} W^2 \quad (52)$$

$$+ \frac{h m^2 \omega_{\text{rfs}}^2}{2 R p_s^4} W^3 - \frac{h \dot{\omega}_{\text{rfs}}}{R \omega_{\text{rfs}}^2} (\underline{\phi} - \phi_s) W$$

$$\Delta(s) = \frac{1}{2\pi R} + \sum_{\ell=1}^{\infty} \frac{1}{\pi R} \cos \frac{\ell(s - s_{\text{gap}})}{R} \quad (53)$$

$$\left\langle \frac{D}{\rho} \right\rangle_{\text{av}} = \frac{1}{\gamma_t^2} \quad (54)$$

$$-K \omega_{\text{rfs}} = \frac{h^2}{R^2 p_s} \left( \frac{1}{\gamma_s^2} - \frac{1}{\gamma_t^2} \right) \quad (55)$$

## References

- [1] K.R. Symon, ANL/APS/TB-28 (1997) (derivation of formulae in this section)
- [2] D.A. Edwards, M.J. Syphers, An Introduction to the Physics of High Energy Accelerators, Wiley (1993) p.129
- [3] H. Bruck, Accélérateurs Circulaires de Particules, Presses Universitaires de France (1966); English translation: LANL Report LA-TR-72-10 (1972)
- [4] E.D. Courant, R.D. Ruth, W.T. Weng, AIP Proc. 127 (1985) p.294
- [5] L.J. Laslett, BNL-7534 (1963) p.324
- [6] A.J. Dragt, AIP Proc. 87 (1982) p.147
- [7] K. Symon, AIP Proc. 249 (1992) p.277
- [8] T. Suzuki, KEK Report 96-10 (1996)

### 2.3.2 Tune Dependence on Momentum and Betatron Amplitudes

*D.A. Edwards, DESY*

*M.J. Syphers, Michigan State U.*

The multipole expansion used here is

$$\Delta B_y + i \Delta B_x = B_0 \sum_{n=1}^{\infty} (b_n + i a_n) (x + iy)^n \quad (1)$$

where  $B_0$  is the field on the reference trajectory, and  $b_n$  ( $a_n$ ) are the normal (skew) multipole coefficients (see also Sec.6.17.1). This form associates  $n$  with a  $2n + 2$  pole magnet.

If a field gradient  $\Delta B' \equiv \partial B_y / \partial x$  is introduced at a location  $s$  over a distance  $ds$  in a closed trajectory of a synchrotron, the bend-plane betatron oscillation tune will be perturbed by an amount

$$d\nu = \frac{1}{4\pi} \oint \beta(s) \frac{\Delta B'(s)}{(B\rho)} ds \quad (2)$$

with an effect of the same magnitude but opposite in sign in the other transverse degree-of-freedom. Then using Eq.(1) the tune change due to a multipole of order  $n$  for a closed path differing from the reference trajectory by  $\Delta x(s)$  will in lowest order be [1]

$$\Delta\nu = \frac{n}{4\pi} \oint \beta(s) \frac{B_0(s)}{(B\rho)} b_n(s) \Delta x(s)^{n-1} ds \quad (3)$$

In the case of a displaced equilibrium orbit due to a momentum offset,  $\Delta x(s)$  in Eq.(3) becomes  $D(s)\Delta p/p_s$  where  $D$  is the dispersion function (Sec.2.1.1). Though a betatron oscillation is not a closed path, Eq.(3) may be used for an estimate with  $\Delta x = a\sqrt{(\beta(s)/\beta_0)} \cos \psi(s)$ .

Applying Eq.(2), the tune shift due to an octopole distribution is

$$\Delta\nu = \frac{3}{8} \langle \beta^2 b_3 \rangle \left( \frac{a^2}{\beta_0} \right) \quad (4)$$

## References

- [1] D.A. Edwards, M.J. Syphers, An Introduction to the Physics of High Energy Accelerators, Wiley (1993)

### 2.3.3 Nonlinear Resonances

*D.A. Edwards, DESY*

*M.J. Syphers, Michigan State U.*

Recursive passage through field imperfections in the magnets of a cyclic accelerator can lead to unwanted resonant behavior resulting in beam instabilities and loss [1]. For example, in lowest nonvanishing order of a perturbative treatment, introduction of a sextupole field can produce instability at betatron oscillation tunes if  $3\nu_x$ ,  $2\nu_x + \nu_y$ ,  $\nu_x + 2\nu_y$  or  $3\nu_y$  are integers. The ubiquity in tune-space of potential resonances for realistic magnet designs leads to reliance on computer simulation for dynamic aperture prediction (Sec.2.3.9 and the relevant material in [2]).

A useful aspect of nonlinear resonances is found in the slow extraction process (Sec.4.12.1). As an illustration, the equations for oscillation development with turn number  $n$  in the neighborhood of a half-integer resonance as given in [3] are

$$\frac{dx}{dn} = p_x \left[ -\frac{1}{2}Q + 2\pi\delta + \frac{3}{8}D(x^2 + p_x^2) \right] \quad (1)$$

$$\frac{dp_x}{dn} = x \left[ -\frac{1}{2}Q - 2\pi\delta - \frac{3}{8}D(x^2 + p_x^2) \right] \quad (2)$$

with  $p_x \equiv \beta x' + \alpha x$ ,  $Q$  is the quadrupole half-integer resonance driving term,  $\delta = |Q/(2\pi)|$  the stop-band width resulting from  $Q$ , and  $D$  is a zero-harmonic  $B'''$  (octupole) term providing a tune dependence with amplitude. For  $B''' > 0$  and stable motion of small amplitude oscillations,  $\delta < 0$  and  $|Q| < 4\pi\delta$  so for extraction the phase stable region will shrink as  $|Q|$  increases. These equations have a first integral (equivalent to a Hamiltonian),

$$\mathcal{H} = [(p_x - u)^2 + x^2 - r^2][(p_x + u)^2 + x^2 - r^2]$$

with  $r^2 = u^2 + v^2$  where  $u$  and  $v$  are the fixed points found by setting  $dx/dn = 0$ ,  $dp_x/dn = 0$  in Eqs.(1,2). Thus the separatrix is formed by two circles the centers of which are the stable fixed points at  $\pm u$ . The intersections are at the unstable fixed points  $\pm v$ , from which particles are “squeezed” out as extraction proceeds.

Though developed to provide external beams for HEP experiments, resonant extraction has application in medical accelerator facilities (Sec.4.12.1), where the third-integer sextupole-driven process is outlined.

### References

- [1] E.D. Courant, H.S. Snyder, Ann. Phys. 1 (1958) 1
- [2] Vol.1, Sec.4.7, LHC Design Report, <http://cdsweb.cern.ch/record/782076>
- [3] D.A. Edwards, M.J. Syphers, An Introduction to the Physics of High Energy Accelerators, Wiley (1993)

### 2.3.4 Synchrobetatron Resonances

*A. Piwinski, DESY*

Synchrobetatron resonances (or satellite resonances) are excited when

$$k\nu_x + \ell\nu_y + m\nu_s = n \quad (1)$$

with  $k$ ,  $\ell$ ,  $m \neq 0$ , and  $n$  integers. These resonances reduce the lifetime of the beams drastically and limit the performance of circular accelerators. Sometimes they are single particle effects, but more often they depend strongly on the bunch current.

The main mechanisms for the excitation of synchrobetatron resonances are dispersions in sextupoles, dispersions in rf cavities, transverse fields in rf cavities, and the beam-beam interaction (Sec.2.5.1) at a crossing angle. The excitation by the natural chromaticity [1, 2], where all sidebands of the integer resonances are excited, is weaker and compensated by sextupoles.

**Dispersion in sextupoles** The changes of  $x'$  and  $y'$  in a sextupole depend on  $k_s[(x_\beta + D_x\delta)^2 - (y_\beta + D_y\delta)^2]$  and on  $k_s(x_\beta + D_x\delta)(y_\beta + D_y\delta)$  where  $\delta = \Delta E/E$ ,  $k_s$  = integrated sextupole strength, and  $D_{x,y}$  = horizontal and vertical dispersion at the sextupoles. These kicks excite the third order betatron resonances  $3\nu_x = n$  and  $\nu_x \pm 2\nu_y = n$ , and the following satellite resonances [3]:

Horizontal satellites

$$2\nu_x \pm \nu_s = n \text{ (excited by } D_x) \\ \nu_x \pm 2\nu_s = n \text{ (excited by } D_x^2 - D_y^2)$$

Vertical satellites

$$2\nu_y \pm \nu_s = n \text{ (excited by } D_x) \\ \nu_y \pm 2\nu_s = n \text{ (excited by } D_x D_y)$$

Satellites of the betatron coupling resonances

$$\nu_x \pm \nu_y \pm \nu_s = n \text{ (excited by } D_y)$$

It is assumed that the synchrotron oscillation energy is much larger than the betatron oscillation energy, so that the synchrotron coordinates are not

changed by the betatron coordinates. For a more precise calculation one has to take into account the path lengthening due to the betatron oscillation [4, 5].

The rise times for the satellites of the betatron coupling resonances are [3]

$$\frac{1}{\tau_{1x,y}} = f_0 \sqrt{\frac{\epsilon_{y,x}}{\epsilon_{x,y}}} \hat{\delta} \times \left| \sum_n k_{sn} D_{yn} \sqrt{\beta_{xn} \beta_{yn}} e^{i(\phi_{xn} \pm \phi_{yn})} \right| \quad (2)$$

where  $\hat{\delta}$  is the energy oscillation amplitude,  $f_0$  is the revolution frequency. For the satellites with a distance of  $\nu_s/2$  from integer and half integer resonances,

$$\frac{1}{\tau_{2x,y}} = f_0 \hat{\delta} \left| \sum_n k_{sn} D_{xn} \beta_{x,yn} e^{2i\phi_{x,yn}} \right| \quad (3)$$

and for satellites with a distance of  $2\nu_s$  from integer resonances,

$$\frac{1}{\tau_{3x}} = \frac{f_0 \hat{\delta}^2}{2\sqrt{\epsilon_x}} \left| \sum_n k_{sn} (D_{xn}^2 - D_{yn}^2) \sqrt{\beta_{xn}} e^{i\phi_{xn}} \right| \quad (4)$$

$$\frac{1}{\tau_{3y}} = \frac{f_0 \hat{\delta}^2}{\sqrt{\epsilon_y}} \left| \sum_n k_{sn} D_{xn} D_{yn} \sqrt{\beta_{yn}} e^{i\phi_{yn}} \right| \quad (5)$$

Note that only the rise times on  $2\nu_{x,y} \mp \nu_s = n$  give an exponential increase. The other rise times depend on the amplitudes and change with increasing amplitude. All these satellites can be compensated by varying the sextupole strength so that the sums in Eqs.(2) to (5) vanish.

**Dispersion in cavities** In a cavity the energy of a particle is changed and its closed orbit is shifted by  $D_x \delta E/E$ . This changes the betatron coordinate by

$$\begin{aligned} \delta x_\beta &= -D_x \frac{\delta E}{E} \\ &= -D_x \frac{eV}{E} \left[ \sin \left( \phi_s + \frac{2\pi z}{\lambda} \right) - \sin \phi_s \right] \end{aligned} \quad (6)$$

with  $V$  = cavity voltage,  $\phi_s$  = synchronous phase,  $\lambda$  = rf wavelength,  $z$  = longitudinal position. A similar equation is obtained for  $\delta x'_\beta$ . Only that part of the energy change which varies with the synchrotron frequency plays a role for a resonance, and not the constant radiation losses.

The change of  $z$  per revolution is [4, 5]

$$\begin{aligned} \delta z &= \oint \frac{1}{\rho} \left( x_\beta + D_x \frac{\Delta E}{E} \right) d\ell \\ &= A_1 x_\beta + A_2 x'_\beta - \alpha_p C \frac{\Delta E}{E} \end{aligned} \quad (7)$$

with

$$A_1 = -\frac{1}{\beta_x} \left[ D_x \sin \mu_x - \tilde{D}_x (1 - \cos \mu_x) + \frac{A_2 \beta'_x}{2} \right]$$

$$A_2 = -D_x (1 - \cos \mu_x) - \tilde{D}_x \sin \mu_x$$

$$\tilde{D}_x = D'_x \beta_x - \frac{1}{2} D_x \beta'_x$$

$\mu_x$  = betatron phase advance,  $\alpha_p$  = momentum compaction factor,  $C$  = circumference.  $x_\beta$ ,  $x'_\beta$ ,  $\beta_x$ ,  $\beta'_x$ ,  $D_x$ , and  $\tilde{D}_x$  refer to the beginning of the revolution.

For linear betatron and synchrotron oscillations an exact solution for the coupled motion can be found. The phases  $\mu_{1,2}$  of the eigenvalues of the revolution matrix, on the resonance  $\nu_x \pm \nu_s = n$ , are [4]

$$\mu_{1,2} = \mu_{x,s} \pm (1 - \cos \mu_x) \sqrt{\frac{-(D_x^2 + \tilde{D}_x^2)}{\alpha_p C \beta_x \sin \mu_x}} \quad (8)$$

Eq.(8) shows that an instability occurs, i.e. the amplitudes grow exponentially, when  $\sin \mu_x > 0$ . This is a difference resonance, since  $\nu_s = \mu_x/2\pi - n$  is always smaller than 0.5.

Nonlinear satellite resonances with  $|m| > 1$  [Eq.(1)] can be investigated approximately. An invariant of motion is [4]

$$\hat{x}_\beta^2 + \frac{\alpha_p C \beta_x}{2\pi m \nu_s} \hat{\delta}^2 = \text{const.} \quad (9)$$

where  $\hat{x}_\beta$  and  $\hat{\delta}$  are the oscillation amplitudes. For most cases the second term in Eq.(9) dominates, and the change of the betatron amplitude is larger than the change of the synchrotron amplitude.

Above transition, for satellite resonances Eq.(1), an instability occurs for a difference resonance. This is opposite to the case of  $x$ - $y$  coupling resonances, for which difference resonances are stable and sum resonances are unstable. Below transition ( $\alpha_p \gamma^2 < 1$ ), one has to replace  $\alpha_p$  by  $(\alpha_p \gamma^2 - 1)/(\gamma^2 - 1)$ ; then satellite resonances behave similarly to the  $x$ - $y$  coupling resonances.

Synchrobetatron resonances due to a dispersion, as satellites of integer resonances, were first observed in NINA [6] and later in SPEAR [7], PETRA [4], DORIS III [8], and LEP [9]. They can be compensated with dispersion bumps.

#### Transverse fields with longitudinal variation

Transverse fields which vary in longitudinal direction inside the bunch can be produced by the accelerating voltage [10] or by the bunch itself [11]. In both cases they are caused by asymmetries of

the cavities or by an off-centered passage of the bunch.

The excitation can be calculated exactly for resonances  $\nu_{x,y} \pm \nu_s = n$ . For horizontal excitation the kick is

$$\delta x' = \frac{e}{p} \int (E_x + vB_y) dt \quad (10)$$

The integral is taken along the path of the particle between two limits where the fields vanish. The kick can then be linearized to yield

$$\delta x' = Gz \quad \text{where} \quad G = \frac{e}{p} \int \frac{\partial E_z}{\partial x} dt \quad (11)$$

The same fields also perturb synchrotron motion. The corresponding energy change which varies linearly with the transverse position of a particle in the bunch is

$$\frac{\delta E}{E} = x \frac{e}{E} \int \frac{\partial E_s}{\partial x} v dt = \frac{v^2}{c^2} Gx \approx Gx \quad (12)$$

Eqs.(11) and (12) describe a linear coupling. On the linear resonance  $\nu_x \pm \nu_s = n$  the phases of the eigenvalues of the revolution matrix are

$$\mu_{1,2} = \mu_{x,s} \pm \frac{G}{2} \sqrt{\frac{\alpha_p C \beta_{xc}}{-\sin \mu_x}} \quad (13)$$

where  $\beta_{xc}$  is the  $\beta$ -function at the cavity. The oscillation amplitudes increase exponentially on a difference resonance. Assuming a constant synchrotron amplitude, the growth rate for the betatron amplitude  $\hat{x}_\beta$  is [4]

$$\frac{d\hat{x}_\beta}{dt} = \frac{1}{2} G \beta_x f_0 \hat{z} \quad (14)$$

where  $\hat{z}$  is the synchrotron amplitude.

These satellites can be compensated with help of orbit bumps in the cavity sections. Since the excitation by a dispersion and the excitation by transverse fields are additive, both mechanisms can amplify or compensate each other depending on the phases of the dispersion and the orbit. Since most of the satellites are current dependent, the compensation is complete only for a specific bunch current.

### Beam-beam interaction with a crossing angle

A particle with longitudinal distance  $z$  from the center of a bunch and a betatron displacement  $x_\beta$  receives a beam-beam transverse kick ( $D_x = 0$ )

$$\delta x'_\beta = f(x_\beta + \phi z) \quad (15)$$

with  $f$  the function describing the beam-beam force and  $2\phi$  the total crossing angle. The synchrotron oscillation is also influenced by the betatron oscillation,  $\delta p_z = \phi \delta p_x$ , and

$$\frac{\delta E}{E} \approx \frac{\delta p}{p} \approx \frac{\delta p_z}{p} \approx \phi \frac{\delta p_x}{p} = \phi \delta x'_\beta = \phi f(x_\beta + \phi z) \quad (16)$$

The complete coupling is described by Eqs.(15) and (16). For small oscillation amplitudes one gets a linear coupling

$$f(x_\beta + \phi z) \approx -\frac{4\pi\xi_x}{\beta_x^*}(x_\beta + \phi z) \quad (17)$$

with

$$\xi_x = \frac{r_0 N_B \beta_x^*}{2\pi\gamma\sigma_{x,\text{eff}}(\sigma_{x,\text{eff}} + \sigma_y)}$$

$\sigma_{x,\text{eff}} = \sqrt{\sigma_x^2 + \phi^2 \sigma_z^2}$ ,  $N_B$  = number of particles per bunch,  $\beta_x^*$  =  $\beta$ -function at the IP,  $\sigma_{z,x,y}$  = rms longitudinal, horizontal and vertical bunch sizes. With linear coupling, the phases  $\mu_{1,2}$  of the eigenvalues are [4]

$$\mu_1 = \mu_x + 2\pi\xi \pm 2\pi\xi\phi \sqrt{\frac{-\alpha_p C}{\beta_x^* \sin \mu_x}} \quad (18)$$

$$\mu_2 = \mu_s - 2\pi\xi\phi^2 \frac{\alpha_p C}{\mu_s \beta_x^*} \pm 2\pi\xi\phi \sqrt{\frac{-\alpha_p C}{\beta_x^* \sin \mu_x}} \quad (19)$$

See also Sec.2.5.1.3.

Computer simulations [12] can be used to determine the strength of higher order satellites. In case of a horizontal crossing all satellites can be excited for which  $k+m$  is even (for antisymmetric space charge force and beam centers colliding head-on). Many resonances are suppressed if the IPs are equidistant. If the phase advances between the IPs are different, e.g. due to machine errors, more resonances can be excited.

Experiments with DORIS I [12] as well as the simulations show that the synchro-betatron resonances excited by the beam-beam interaction with a crossing angle reduce mainly the lifetime and do not affect seriously the beam size or the luminosity. Four parameters determine the reduction of the lifetime: (i) the normalized crossing angle  $\phi\sigma_z/\sigma_x$ , (ii) the beam-beam parameter  $\xi$ , (iii) the synchrotron tune  $\nu_s$ , and (iv) the physical or dynamical aperture.

### References

- [1] Y. Orlov, Sov. Phys. JETP, V5 (1957) 45
- [2] K.W. Robinson, CEA-54 (1958)

- [3] A. Piwinski, DESY 93-187 (1993)
- [4] A. Piwinski, 11th Int. Conf. on High-Energy Acc. (1980) p.638
- [5] A. Piwinski, A. Wrulich, DESY 76/07 (1976)
- [6] M.C. Crowley-Milling, I.I. Rabinowitz, PAC 77, p.1052
- [7] SPEAR Group, PAC 75, p.1366
- [8] H. Nesemann, 15th Int. Conf. on High Energy Acc. (1992) p.549
- [9] J.M. Jowett et al, EPAC 90, p.1612
- [10] N.A. Vinokurov et al, 10th Int. Conf. on High Energy Acc. (1977) p.272
- [11] R.M. Sundelin, PAC 79, p.3604
- [12] A. Piwinski, PAC 77, p.1408

### 2.3.5 Taylor Map, Hénon Map, Standard Map

*A. Dragt, U. Maryland*

**Definition** A *transfer map*  $\mathcal{M}$  is a *relation* that sends *initial conditions*  $\vec{z}^i$  to *final conditions*  $\vec{z}^f$ , which we write symbolically in the form

$$\vec{z}^f = \mathcal{M}\vec{z}^i \quad (1)$$

A Taylor map is a representation of  $\mathcal{M}$  that presents final conditions as a Taylor series in the initial conditions. Let  $\vec{\zeta}$  refer to deviation coordinates about some reference orbit. Then  $\mathcal{M}$  has a Taylor representation [1]

$$\begin{aligned} \zeta_j^f = & \sum_k R_{jk} \zeta_k^i + \sum_{kl} T_{jkl} \zeta_k^i \zeta_l^i \\ & + \sum_{k\ell m} U_{jk\ell m} \zeta_k^i \zeta_\ell^i \zeta_m^i + \dots \end{aligned} \quad (2)$$

The number of monomials of degree 0 through  $D$  in  $N$  variables is  $(D+N)!/(D!N!)$ . Therefore, if (2) is truncated beyond terms of degree  $D$ , the number of coefficients required to describe a *truncated Taylor map* is  $N(D+N)!/(D!N!)$ . In the Hamiltonian case the coefficients are interconnected by nonlinear relations arising from the symplectic condition, and are thus not independent. Extensive tools have been developed for manipulating truncated power series (Sec.2.3.7).

**Existence of Taylor maps** The equations of motion for a charged particle in an EM field, including the average (but not the stochastic) effect of synchrotron radiation, can be written in the form

$$\frac{dz_j}{ds} = f_j(\vec{z}; s) = f_j(z_1, \dots, z_N; s), \quad j = 1, N \quad (3)$$

Given initial conditions  $\vec{z}^i$  at reference position  $s^i$ , (3) can be integrated to yield unique final conditions  $\vec{z}^f$  at position  $s^f$ . We write this relation between  $\vec{z}^i$  and  $\vec{z}^f$  as (1). Integrating backward in time, the map  $\mathcal{M}$  arising from (3) is always invertible,  $\vec{z}^i = \mathcal{M}^{-1}\vec{z}^f$ . Theorem of Poincaré: If the  $f_j(\vec{z}; \lambda; s)$  are *analytic* (have convergent Taylor series) in  $s$  and  $\vec{z}$ , then  $\vec{z}^f$  is analytic in  $\vec{z}^i$ . For charged particle motion, the  $f_j$  involve electromagnetic fields obtained from the Maxwell equations and the conditions for Poincaré's theorem are met. Therefore the expansion (2) always exists and converges for sufficiently small  $\vec{\zeta}^i$ . Analytic dependence on parameters  $\lambda$ , e.g. field strength, can be treated similarly.

**Calculation of Taylor maps** In the approximation that extended fringe-field effects are neglected, there are  $s$ -independent  $f_j$  for each idealized beamline element. The design orbit for each such element is then known, and the Taylor map about the design orbit can be found by letting the associated Lie transformation (Sec.2.3.7) act on  $\vec{\zeta}$ . If  $s$ -dependence is taken into account, as is necessary for the accurate treatment of fringe-field effects and wigglers/undulators, one can integrate (3) to find a design orbit and integrate the associated *complete variational equations* to find the Taylor expansion (2) of  $\mathcal{M}$  about this design orbit. There is a machine-implementable algorithm that will generate the complete variational equations to any desired order. If there is a Hamiltonian formulation for the equations of motion, one may also use a symplectic integrator (Sec.2.3.8) to find the Taylor coefficients or may directly integrate equations for the Lie generators of  $\mathcal{M}$ . The equations of motion for the Lie generators are known through 8<sup>th</sup> order. This approach, which exploits the symplectic condition, involves integrating roughly 3-times fewer equations than those required to find the Taylor coefficients [2].

**Manipulation of Taylor maps** Multiplication (*concatenation*): If maps for individual elements are known in Taylor form, these series may be substituted into each other to find a net Taylor map. Factorization: Maps in Taylor form may be factored into the product of nonsymplectic and symplectic parts, each in Lie form and unique. Inversion: Taylor maps may be inverted by series reversion implemented by inverting  $R$  followed by recursive back substitution. Normal forms: Let  $\mathcal{M}$  be a given map. One seeks *normalizing* maps

$\mathcal{A}$  such that the map  $\mathcal{N} = \mathcal{A}\mathcal{M}\mathcal{A}^{-1}$  has a particularly simple form called a *normal* form.

**Applications of Taylor maps** Maps can be used to look at aberration (nonlinear) content, to monitor a design process, and to replace and speed up element-by-element tracking. To look at aberration content one converts the Taylor map to a Lie map, in which each Lie coefficient corresponds to a unique and readily identifiable aberration (Sec.2.3.6). For storage rings, these aberrations can also be specified in an action-angle *resonance* basis. To monitor a design process one examines the aberration coefficients at each step to ascertain hoped-for improvements. Designs and design codes can be compared and cross-checked by comparing aberration strengths. In a beamline, one or more segments may be of special interest. For example, in a final focus system,  $s^f$  is chosen to be the IP. To avoid having the map coefficients depend on the choice of  $s^f$ , one can factor the map as a linear map from  $s^f$  to the IP, followed by a nonlinear map from the IP to itself. For a circular machine the one-turn map can be found by concatenation and then brought to normal form  $\mathcal{N}$ .  $\mathcal{N}$  contains all information about tunes, phase slip, chromaticities, and anharmonicities.  $\mathcal{A}$  contains all information about lattice (betatron) functions. All couplings are automatically included. The  $\delta$ -dependent chromatic terms, which are isolated by this procedure, often need special attention because the longitudinal emittance is typically orders of magnitude larger than the transverse. It is important to understand how these  $\delta$ -dependent terms arise/cancel in beam-line sections and to devise schemes that minimize this dependence. Together  $\mathcal{A}$  and the amplitude dependent terms (anharmonicities) and chromaticities determine the tune footprint for any given beam distribution. The tune footprint can also be minimized.

When due attention is paid to the symplectic condition, experience shows that tracking with a high-order one-turn map, or a few high-order *lumped* maps, gives the same results (comparable dynamic aperture) as element-by-element tracking. Use of the one-turn map or a few lumped maps can be at least an order of magnitude faster, and realistic maps based on 3-D field data and including all fringe-field and high-order multipole effects, rather than the approximate thin-element *kick* maps employed in element-by-element tracking, can be used without sacrificing speed. The

symplectic condition can be enforced by use of mixed-variable generating functions or Cremona maps. Both can be computed from the truncated Taylor map. Among generating functions, that of Poincaré seems optimal. In summary, the use of Taylor maps, expressed in different coordinate systems, and their transformation to Lie and normal forms, have become standard tools of contemporary optical system design. These maps provide insight into the presence, nature, origin, and effect of aberrations/nonlinearities, and ways to compensate/exploit them [2, 3].

**Hénon map and standard map** [4] Requiring that the general quadratic Taylor map in two variables be symplectic yields the much studied *Hénon* map. It can be written as

$$\begin{aligned} x_{n+1} &= x_n \cos \mu + (p_n + Sx_n^2) \sin \mu \\ p_{n+1} &= -p_n \sin \mu + (p_n + Sx_n^2) \cos \mu \end{aligned} \quad (4)$$

In Lie form (Sec.2.3.6), it is given by  $\mathcal{M} = \mathcal{R}(\mu) \exp(:Sx^3/3:)$ . This map corresponds to a clockwise phase-space rotation by angle  $\mu$  (phase advance), followed by a thin sextupole kick of strength  $S$ . Although simple in appearance, when viewed as a one-turn map it is complicated under iteration and has a tune-dependent dynamic aperture (Sec.2.3.9). Thus, it provides a simple example that can be studied to gain insight into the general question of dynamic apertures. It has been generalized to higher dimensions [5].

The simplest synchrotron motion map, associated with a thin rf cavity kick, can be written as (Sec.2.1.2)

$$\begin{aligned} \delta_{n+1} &= \delta_n + \frac{eV}{\beta^2 E} (\sin \phi_n - \sin \phi_s) \\ \phi_{n+1} &= \phi_n + 2\pi h\eta(\delta_{n+1}) \delta_{n+1} \end{aligned} \quad (5)$$

If one defines  $y \equiv h|\eta(\delta)|\delta$ ,  $x \equiv \phi/2\pi$ ,  $\nu_s$  = the synchrotron tune, then (5) becomes for stationary-bucket synchrotron motion the much studied *standard map*:

$$\begin{aligned} y_{n+1} &= y_n - 2\pi\nu_s^2 \sin 2\pi x_n \\ x_{n+1} &= \text{mod}(x_n + y_{n+1}, 1) \end{aligned} \quad (6)$$

The Lie form for the standard map is  $\mathcal{M}_s = \exp(\nu_s^2 \cos 2\pi x) \exp(-y^2/2)$ . This map exhibits chaotic behavior that becomes large-scale rather quickly as  $\nu_s$  is varied.

## References

- [1] K. Brown, SLAC-75-REV-4 (1982) and SLAC-R-95-462 (1995)

- [2] A. Dragt, Lie Methods for Nonlinear Dynamics with Applications to Accelerator Physics (2010); [www.physics.umd.edu/dsat/](http://www.physics.umd.edu/dsat/)
- [3] M. Berz, Advances in Imaging and Electron Physics 108, P. Hawkes, Edit., Academic (1999)
- [4] Google “Henon Map” and “Standard Map”
- [5] J. Moser, Mathematische Zeitschrift 216, 417 (1994)

### 2.3.6 Lie Algebraic Methods

*A. Dragt, U. Maryland*

Map and Lie methods provide a framework for abstract concepts and related computational algorithms. They provide a powerful extension of linear concepts and algorithms into the nonlinear regime [1, 2, 3, 4, 5].

**Definitions** Let  $f$  and  $g$  be functions of the phase-space variables  $z = (q_1, p_1, q_2, p_2, \dots)$ . Although results are general, we will assume that phase space is 6-D.

*Poisson bracket:*

$$[f, g] = \sum_{j=1}^3 \left( \frac{\partial f}{\partial q_j} \frac{\partial g}{\partial p_j} - \frac{\partial f}{\partial p_j} \frac{\partial g}{\partial q_j} \right) \quad (1)$$

*Fundamental Poisson brackets:*

$$[z_a, z_b] = J_{ab}, \quad a, b = 1, 6 \text{ where} \quad (2)$$

$$J = \begin{bmatrix} J_2 & 0 & 0 \\ 0 & J_2 & 0 \\ 0 & 0 & J_2 \end{bmatrix}, \quad J_2 = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \quad (3)$$

*Symplectic matrices:* Matrices  $N$  that satisfy  $N^T J N = J$  are called symplectic.

*Symplectic maps:* Given a transfer map  $\mathcal{M}$ , form the Jacobi matrix  $M(z^i)$ ,

$$M_{ab}(z^i) = \partial z_a^f / \partial z_b^i \quad (4)$$

$\mathcal{M}$  is called symplectic if  $M(z^i)$  is symplectic for all  $z^i$ . Truncating a Taylor map generally violates the symplectic condition.

*Lie operators:* Associated with any function  $f(z, t)$  of phase-space variables  $z$  and time  $t$  is a differential *Lie operator* denoted by  $:f:$ ,

$$:f: = \sum_{j=1}^3 \left( \frac{\partial f}{\partial q_j} \frac{\partial}{\partial p_j} - \frac{\partial f}{\partial p_j} \frac{\partial}{\partial q_j} \right) \quad (5)$$

More precisely,  $:f:$  is a *Hamiltonian Lie operator*. Let  $x = (x_1 \dots x_N)$  and let  $g_1(x) \dots g_N(x)$  be a collection of  $N$  functions of  $x$ . A general Lie operator (or *vector field*)  $\mathcal{L}_g$  is defined by

$$\mathcal{L}_g = \sum_{b=1}^N g_b(x) (\partial/\partial x_b) \quad (6)$$

Hamiltonian Lie operators are when  $N$  is even and  $g_b(z) = \sum_a (\partial f/\partial z_a) J_{ab}$ . Non-Hamiltonian vector fields are used for dissipative effects (e.g. synchrotron radiation).

*Lie transformations:* Associated with  $:f:$  or  $f$  is the Lie transformation denoted by  $\exp(:f:)$  and defined by

$$\exp(:f:) = e^{:f:} = \sum_{n=0}^{\infty} :f:^n / n! \quad (7)$$

More precisely,  $\exp(:f:)$  is a *Hamiltonian Lie transformation*. Operators  $\exp(\mathcal{L}_g)$  are also called Lie transformations. The quantities  $f$ , or  $:f:$ , or  $\mathcal{L}_g$  are called *Lie generators*.

**General facts/theorems** Symplectic matrices form a Lie group, called  $Sp(2n, \mathbb{R})$ , of dimension  $n(2n+1)$  where  $2n$  is the phase-space dimension. Symplectic maps form an infinite dimensional Lie group. If  $\mathcal{M}$  arises from integrating Hamiltonian differential equations, it will be symplectic, and vice versa. For any  $g(z, t)$ , Lie operators have the properties

$$\begin{aligned} :f:^0 g &= g, & :f:g &= [f, g], & :f:^2 g &= [f, [f, g]], \dots \\ \exp(:f:)g &= g + [f, g] + [f, [f, g]]/2! + \dots \end{aligned} \quad (8)$$

The sum on the right side of (8) is called a *Lie series*.

Suppose  $f(z^i, t)$  is any function of  $z^i$  and possibly the time  $t$ . Define

$$z^f = \exp(:f:)z^i \quad (9)$$

Then  $z^f$  and  $z^i$  are related by a *symplectic map*  $\mathcal{M}$ ,

$$z^f = \mathcal{M}z^i \quad \text{with} \quad \mathcal{M} = \exp(:f:) \quad (10)$$

That is, Lie transformations provide an endless supply of symplectic maps. Conversely, suppose  $\mathcal{M}$  is an analytic symplectic map having a Taylor expansion of  $z^f$  in terms of  $z^i$ . Then  $\mathcal{M}$  has the Dragt-Finn factored product representation

$$\begin{aligned} \mathcal{M} = \exp(:f_1:) \exp(:f_2^c:) \exp(:f_2^a:) \\ \exp(:f_3:) \exp(:f_4:) \dots \end{aligned} \quad (11)$$

Here the  $f_m$  are *homogeneous polynomials* of degree  $m$ . The quadratic polynomials  $f_2^{c,a}$  are of the form

$$f_2^{c,a} = -\frac{1}{2} \sum_{de} S_{de}^{c,a} z_d z_e \quad (12)$$

where  $S^c$  and  $S^a$  are any symmetric matrices that *commute* or *anticommute* with  $J$ , respectively. We remark that any Taylor map can be uniquely

factored into a symplectic (Hamiltonian) and non-symplectic (non-Hamiltonian) part [4]. The symplectic part then has the factorization (11) and the nonsymplectic part has an analogous Lie factorization in terms of non-Hamiltonian vector fields. Finally, unlike the Taylor representation, the product (11) can be truncated at any stage and the result will still be a symplectic map.

Introduce the shorthand notation

$$\mathcal{R}_f = \exp(:f_2^c:) \exp(:f_2^a:) \quad (13)$$

Each factor in (11) plays a definite role. The factor  $\exp(:f_1:)$  produces phase-space translations and  $\mathcal{R}_f$  produces the linear part of the Taylor map described by the matrix  $R$ . See Sec.2.2.1. The factor  $\exp(:f_3:)$  produces the quadratic nonlinear part described by  $T$  and also contributes to still higher-order parts described by  $U, \dots$ ; the factor  $\exp(:f_4:)$  produces the cubic nonlinear part described by  $U$  and also contributes to still higher-order parts, etc. Indeed, there is a well-defined procedure for passing between Taylor series and Lie product representations of symplectic maps. To reproduce the Taylor coefficients for a symplectic map  $\mathcal{M}$  through terms of degree  $D$ , it is necessary to specify the generators  $f_1, f_2 \dots f_{D+1}$ . In the case of  $N$  variables this requires  $[-1 + (D+1+N)!/(D+1)!N!]$  coefficients, which is considerably fewer than the number of Taylor coefficients. In summary,  $\exp(:f_1:)$  and  $\mathcal{R}_f$  describe translations and linear transformations, and the remaining factors describe nonlinear effects. And this is done in an optimal way since any set of  $f_m$  produces a unique symplectic map, and any analytic symplectic map can be expressed as (11) with the  $f_m$  uniquely determined. Later, after introducing suitable coordinates specific to accelerator physics applications, we will summarize some of the simpler Lie maps that are applicable to various idealized beam-line elements.

**Lie algebraic facts** There are deep properties of the exponential function which facilitate the manipulation of maps in Lie form.

**Combining exponents:** Let  $A$  and  $B$  be any two square matrices of the same dimension. Form matrices  $\exp(sA)$  and  $\exp(tB)$  where  $s$  and  $t$  are parameters. For  $s$  and  $t$  sufficiently small, a matrix  $C$  that satisfies  $\exp(sA)\exp(tB) = \exp(C)$  exists, is unique, and is a sum of elements formed only from  $A$  and  $B$  and their multiple commuta-

tors, for which we use the notation  $\{A, B\}$ , etc.,

$$\begin{aligned} C(s, t) &= sA + tB + (1/2)\{A, B\} \\ &+ \frac{1}{12}s^2t\{A, \{A, B\}\} + \frac{1}{12}st^2\{B, \{B, A\}\} \\ &- (1/24)s^2t^2\{A, \{B, \{A, B\}\}\} + \dots \end{aligned} \quad (14)$$

That  $C$  depends only on commutators is the *Baker-Campbell-Hausdorff* (BCH) theorem. The series (14) is called the BCH series. In the setting of Lie transformations,  $\exp(:f:) \exp(:g:) = \exp(:h:)$  with

$$\begin{aligned} h &= f + g + (1/2)[f, g] + (1/12)[f, [f, g]] \\ &+ \frac{1}{12}[g, [g, f]] - \frac{1}{24}[f, [g, [f, g]]] + \dots \end{aligned} \quad (15)$$

**Inversion formulas:** It follows from the BCH formula that the inverses of (10),(11) are

$$\begin{aligned} \mathcal{M}^{-1} &= \exp(-:f:) \\ \mathcal{M}^{-1} &= \dots e^{-:f_4:}e^{-:f_3:}e^{-:f_2^a:}e^{-:f_2^c:}e^{-:f_1:} \end{aligned} \quad (16)$$

**Map multiplication:** Define the matrix  $R^f$  associated with  $\mathcal{R}_f$  by the rule

$$\mathcal{R}_f z_a = \sum_b R_{ab}^f z_b \quad (17)$$

Two symplectic maps  $\mathcal{M}_f = \mathcal{R}_f \exp(:f_3:) \exp(:f_4:) \dots$  and  $\mathcal{M}_g = \mathcal{R}_g \exp(:g_3:) \exp(:g_4:) \dots$  can be multiplied (concatenated) to find  $\mathcal{M}_h = \mathcal{M}_f \mathcal{M}_g$  with  $\mathcal{M}_h = \mathcal{R}_h \exp(:h_3:) \exp(:h_4:) \dots$ . Use of the BCH series gives the results

$$\begin{aligned} R^h &= R^g R^f \\ h_3 &= f_3^{\text{tr}} + g_3 \\ h_4 &= f_4^{\text{tr}} + g_4 + (1/2)[f_3^{\text{tr}}, g_3] \\ h_5 &= f_5^{\text{tr}} + g_5 - [g_3, f_4^{\text{tr}}] - (1/6):f_3^{\text{tr}}:^2 g_3 \\ &\quad + (1/3):g_3:^2 f_3^{\text{tr}}, \text{ etc.} \end{aligned} \quad (18)$$

Here  $f_m^{\text{tr}}$  denotes the *transformed* polynomial  $f_m^{\text{tr}}(z) = f_m[(R^g)^{-1}z]$ . Results are known through  $h_8$  and when  $\mathcal{M}_f$  and  $\mathcal{M}_g$  also contain  $f_1$  and  $g_1$  factors. Map multiplication in Lie form is much faster than in Taylor form. The price to be paid for this speed is that formulas of the kind just quoted have to be derived and coded in advance for any specified order whereas substitution of Taylor series into Taylor series can be done using a machine-implementable algorithm that can be initiated to any specified order at run time. Concatenation formulas are used to find the net map for successive beam-line elements and for general map manipulation including the conversion of  $\mathcal{M}^{-1}$  as given by (16) into standard factored product form. In a circular machine, they are also used to find the one-turn (or multiple-turn) map.

**Canonical coordinates** Here we introduce canonical coordinates that are useful for accelerator physics applications. We will work with a full 6-D phase space.

*Unscaled coordinates:*  $z = (x, p_x, y, p_y, \tilde{t}, \tilde{p}_t)$  where  $\tilde{p}_t = p_t - p_t^0$  with  $p_t = -(m^2 c^4 + p^2 c^2)^{1/2} = -\gamma m c^2$  and  $p_t^0 = -[m^2 c^4 + (p^0 c)^2]^{1/2}$  and  $p^0$  the design momentum. The coordinate  $\tilde{t}$  is time of flight relative to the synchronous particle.

*Dimensionless scaled coordinates:* Here  $\ell$  is a convenient scale length.

$$\begin{aligned} X &= x/\ell, P_x = p_x/p^0 \\ Y &= y/\ell, P_y = p_y/p^0 \\ \tau &= c\tilde{t}/\ell, P_\tau = \tilde{p}_t/(p^0 c) \end{aligned} \quad (19)$$

*Relation to noncanonical coordinates:* With  $\beta_0 = v^0/c = -p^0 c/p_t^0$ ,  $P_\tau$  and the momentum deviation  $\delta = (p - p^0)/p^0$  are related by the formulae

$$\begin{aligned} P_\tau &= -(1/\beta_0)\{[1 + (2\delta + \delta^2)\beta_0^2]^{1/2} - 1\} \\ &= -\beta_0\delta + (\delta^2/2)(\beta_0^3 - \beta_0) + \dots \end{aligned} \quad (20)$$

$$\begin{aligned} \delta &= [1 - 2P_\tau/\beta_0 + P_\tau^2]^{1/2} - 1 \\ &= -P_\tau/\beta_0 + (P_\tau^2/2)(1 - \beta_0^{-2}) + \dots \end{aligned} \quad (21)$$

With  $s = z$  as the independent variable and  $(X, P_x, Y, P_y, \tau, P_\tau)$  as dependent variables, the Hamiltonian for motion in a drift space is

$$H = -\frac{1}{\ell}\left[\left(1 - \frac{2P_\tau}{\beta_0} + P_\tau^2 - P_x^2 - P_y^2\right)^{\frac{1}{2}} + \frac{P_\tau}{\beta_0} - \frac{1}{\beta_0^2}\right] \quad (22)$$

Using  $H$ , we find that  $x' = dx/dz = \ell X' = \ell(\partial H/\partial P_x)$  and hence

$$\begin{aligned} x' &= P_x \left(1 - \frac{2P_\tau}{\beta_0} + P_\tau^2 - P_x^2 - P_y^2\right)^{-\frac{1}{2}} \\ &= P_x + \frac{P_x P_\tau}{\beta_0} + \frac{P_x}{2} \left[P_\tau^2 \left(\frac{3}{\beta_0^2} - 1\right) + P_x^2 + P_y^2\right] + \dots \end{aligned}$$

The TRANSPORT variable  $x'$  equals  $P_x$  in lowest order, but has second-order chromatic and higher-order geometric and chromatic differences. The variables  $X, x'$  do not form a canonical pair,  $[X, x'] \neq 1$ . Consequently, the use of TRANSPORT variables makes it difficult to verify or exploit the symplectic condition.

**Computation of maps** For each beamline element,  $\mathcal{M}$  obeys the equation of motion  $d\mathcal{M}/ds = \mathcal{M} : H$ : with the initial condition  $\mathcal{M}(0) = \mathcal{I}$ .

*Autonomous case:* In the approximation/idealization that the Hamiltonian  $H$  of a

straight beamline element (length  $L$ ) is independent of  $s$ , the map for the body of the element is given in *single exponent* form

$$\mathcal{M} = \exp(-L:H:) \quad (23)$$

(There is an analogous result for curved elements.) Here  $H$  is assumed to be expanded in homogeneous polynomials,

$$H = H_2 + H_3 + H_4 + \dots \quad (24)$$

Table 1 gives a few examples of single-exponent maps in lowest order for some simple beam-line elements/transformations.

Table 1: Example Lie maps for some ideal beam-line elements.

| Beamline element                                     | Lie Map   |
|--|---|
| Drift length $L$<br>(lowest order)                   | $\exp[-\frac{L}{2\ell}:P_x^2 + P_y^2 + \frac{P_\tau^2}{(\beta_0\gamma_0)^2}:]$  |
| Thin quad. $Q$                                       | $\exp(-\frac{Q}{2}:X^2 - Y^2:)$   |
| Thin sext. $S$                                       | $\exp(-\frac{S}{3}:X^3 - 3XY^2:)$   |
| Kick $\Delta P_x = f(X)$                             | $\exp[: \int_0^X f(x)dx:]$  |
| $x$ -deflection $d$                                  | $\exp(d:X:)$  |
| Thick quad. $k, L$                                   | $\exp(-L:H:) \text{ where } H = (\frac{1}{2\ell})[(P_x^2 + P_y^2 + \frac{P_\tau^2}{\beta_0^2\gamma_0^2}) + k^2(X^2 - Y^2)]$ |
| $x$ -phase advance $\theta$                          | $\exp(-\frac{\theta}{2}:P_x^2 + X^2:)$  |
| $x$ -phase advance<br>with Courant-Snyder parameters | $\exp(-\frac{\theta}{2}:\beta P_x^2 + 2\alpha X P_x + \gamma X^2:)$   |

A *single exponent* Lie map can be converted to Taylor form by letting it act on  $z$ . In general many terms need to be retained in the resulting Lie series and the coefficients  $1/n!$  must be relied on to yield convergence. To convert a single exponent form to a factored product form, one way is to convert the Taylor series just described to factored product Lie form. Other ways, faster, and better convergent for large  $LH$ , involve scaling, splitting, and squaring methods.

*Nonautonomous case:* When the Hamiltonian is not constant (as occurs in reality due to fringe fields and field nonuniformity including wigglers), the transfer map is first written in the reverse factorized form

$$\mathcal{M} = \dots \exp(:f_4:) \exp(:f_3:)\mathcal{R}. \quad (25)$$

Next, let  $S$  be the symmetric matrix with  $H_2 = (1/2) \sum_{ab} S_{ab} z_a z_b$ . Then the matrix  $R$  that describes  $\mathcal{R}$  obeys

$$dR/ds = JSR, \quad \text{initial cond. } R(0) = I \quad (26)$$

Similarly, the  $f_m$  obey

$$\begin{aligned} df_3/ds &= -H_3^{\text{int}} \\ df_4/ds &= -H_4^{\text{int}} + (1/2):f_3:(-H_3^{\text{int}}) \\ df_5/ds &= -H_5^{\text{int}} + :f_3:(-H_4^{\text{int}}) \\ &\quad +(1/3):f_3:^2(-H_3^{\text{int}}), \text{ etc.} \end{aligned} \quad (27)$$

with the initial conditions  $f_m(0) = 0$ . Results are known through  $f_8$ . Here the *interaction picture* quantities  $H_m^{\text{int}}$  are the polynomials  $H_m^{\text{int}}(z) = H_m(Rz)$ . Eqs.(26) and (27) are to be integrated numerically. The work involved in this Lie approach is considerably less than that involved in integrating the complete variational equations for the corresponding Taylor map. Again, the price to pay for this speed advantage is that formulae of this kind have to be derived and coded in advance for any specified order whereas there is a machine-implementable algorithm that can be initiated at run time to generate the complete variational equations to any specified order. (Both in this case and for concatenation formulas it is pleasant that there are two complementary approaches for checking each other.) Once  $R$  and the  $f_m$  are determined, (25) can be converted to standard factored product form by concatenation.

**Realistic maps** In the discussion above it is assumed that the quantities  $H_m$  in (24) are known. Since the Hamiltonian for charged-particle motion in an EM field involves the vector potential  $\mathbf{A}$ , this assumption is equivalent to assuming high-order Taylor expansions of  $\mathbf{A}$  are known. Equivalently, high-order derivatives of the field are assumed known. Realistic field data can be provided on a grid with the aid of various 3-D finite element codes, sometimes spot checked against measured data. But the computation of high-order transfer maps based on this data appears to pose an insurmountable problem: the direct calculation of high derivatives based only on grid data is intolerably sensitive to noise (due to truncation or round-off) in the grid data. This problem can be solved by the use of surface methods (Sec.2.2.2). The effect of numerical noise can be overcome by fitting field data on a bounding surface far from the beam axis and continuing inward using the Maxwell equations. While the process of differentiation serves to amplify the effect

of numerical noise, the process of continuing inward using the Maxwell equations is *smoothing*. This smoothing is related to the fact that harmonic functions take their extrema on boundaries. When using surface methods, all fits are made on such boundaries. Therefore, if these fits are accurate, interior data based on these fits will be even more accurate. Thus, with the use of surface methods, it is possible to compute for the first time accurate transfer maps for realistic beam-line elements. [4, 6, 7].

**Applications of maps** There are four well-developed uses for maps: analysis, optimization/fitting, tracking, and the treatment of moments including emittance questions.

**Analysis (single-pass systems):** Write the total transfer map  $\mathcal{M}$  in the form  $\mathcal{M} = \mathcal{L}\mathcal{T}$  where  $\mathcal{L}$  is a map (often linear) describing the desired behavior for the system and  $\mathcal{T}$  describes aberrations. Table 2 gives examples of aberration maps of a spot forming or final focus system. Suppose the system is made of two sections  $\mathcal{M}_{1,2}$  and a corrector  $\mathcal{C}$  is placed between them. Then  $\mathcal{T}$  is replaced by  $\mathcal{T}'$  with  $\mathcal{T}' = \mathcal{T}\mathcal{M}_2^{-1}\mathcal{C}\mathcal{M}_2$ . Analogous formulae apply for the case of several correctors at several locations. The correctors and their locations are selected so that  $\mathcal{T}'$  is free of offensive generators.

Table 2: Example aberration maps.

| Aberration   | Aberration map $\mathcal{T}$             |
|--|--|
| Chrom. horiz. steering<br>(dispersion), strength $e$ | $\exp(e:P_x P_\tau:)$                    |
| Chrom. horiz. focusing<br>strength $e$               | $\exp(e/2:P_x^2 P_\tau:)$                |
| Third-order spherical                                | $\exp(:aP_x^4 + bP_x^2 P_y^2 + cP_y^4:)$ |

**Analysis (circulating systems):** Here, *normal form* analysis is useful [2, 3, 5]. Let  $\mathcal{M}$  be the one-turn map. Its normal form is  $\mathcal{N} = \mathcal{A}\mathcal{M}\mathcal{A}^{-1}$  where  $\mathcal{A}$  is a symplectic map selected to make  $\mathcal{N}$  as simple as possible. In case of a static map (no rf),  $\mathcal{A}$  can be selected so that  $\mathcal{N} = \exp(-:h:)$  with

$$\begin{aligned} h &= (w_x + w'_x P_\tau + w''_x P_\tau^2)h_x \\ &\quad + (w_y + w'_y P_\tau + w''_y P_\tau^2)h_y \\ &\quad + ah_x^2 + bh_x h_y + ch_y^2 + dP_\tau^2 \\ &\quad + eP_\tau^3 + fP_\tau^4 + \dots \end{aligned} \quad (28)$$

Here  $h_x = (X^2 + P_x^2)/2$ ,  $h_y = (Y^2 + P_y^2)/2$ . In principle (28) can be extended to arbitrary order ( $\sim 12^{\text{th}}$  order in practice) providing the tunes are not resonant to that order. Coefficients in (28):

- $w_x, w_y$  give the “tunes” ( $= w/2\pi$ ) of the on-energy closed orbit,
- $w'_x, w'_y$  are first-order chromaticities,
- $w''_x, w''_y$  are second-order chromaticities,
- $a, b, c$  are “anharmonicities” i.e. how tunes depend on betatron amplitudes.
- $d, e, f$  are linear and higher order “phase-slip” terms,

$\mathcal{N}$  sends circles in the  $X, P_x$  and  $Y, P_y$  planes into themselves, i.e. it describes idealized betatron motion.  $\mathcal{A}$  gives complete information about all linear and nonlinear Courant-Snyder parameters, lattice functions (including nonlinear distortion functions) and their chromatic behavior, the closed orbit, and coupling effects. Thus,  $\mathcal{A}$  and  $\mathcal{A}^{-1}$  describe the transformation between the actual and the idealized betatron motion.

The quantities  $I^{x,y} = \mathcal{A}^{-1}h_{x,y}$  are invariant under the action of  $\mathcal{M}$ , i.e.  $\mathcal{M}I^{x,y} = I^{x,y}$ , and generalize Courant-Snyder invariants to the nonlinear case.

Applying map  $\mathcal{A}^{-1}$  to turn-by-turn tracking data, the transformed data lie on circles in the  $X, P_x$  and  $Y, P_y$  planes to the extent that the normal form process succeeds. This presentation of tracking data provides a test for the existence of KAM tori.

Let  $\mathcal{A}_1$  and  $\mathcal{A}_2$  be the factors in  $\mathcal{A}$  that remove closed-orbit offsets and bring the linear part of the remaining map to normal form, respectively. Write the partially normalized map as

$$\begin{aligned}\mathcal{M}' &= (\mathcal{A}_2\mathcal{A}_1)\mathcal{M}(\mathcal{A}_2\mathcal{A}_1)^{-1} \\ &= \mathcal{R}'_2 \exp(:g_3:) \exp(:g_4:) \cdots\end{aligned}$$

$$\mathcal{R}'_2 = \exp(w_x:h_x: + w_y:h_y: + d:P_\tau^2:)$$

The generators  $g_3, g_4, \dots$  can be expanded in eigenvectors of the Lie operators  $:h_x:$  and  $:h_y:$ . These eigenvectors form the *static resonance basis*. When such an expansion is made, the coefficients of the various resonance basis vectors are the *resonance driving terms* in the map  $\mathcal{M}$ .

Methods of this section can be generalized to time dependent *dynamic* maps (powered rf cavities). All synchrotron oscillation effects are then taken into account including synchrobetatron coupling and resonance effects.

*Optimization/Fitting:* Procedures can be executed in which various beam-line parameters or other parameters are varied to drive selected transfer map parameters, or quantities computed from them, to specified values. This includes fitting tunes, anharmonicities, chromaticities, dispersions, and other lattice functions. It also includes focusing and aberration correction and the construction of high-order achromats. Because Lie generators are independent, while Taylor coefficients are not, there is less likelihood of demanding the impossible [3].

*Tracking:* To compute single-particle transport, it is necessary to evaluate  $\mathcal{M}z^i$  with  $\mathcal{M}$  given by (11). The effect of  $\exp(:f_1:)$  is to produce a phase-space translation, and the  $\exp(:f_2:)$  factors produce the linear transformation (13). The effect of the remaining product of factors  $\mathcal{N} = \exp(:f_3:) \exp(:f_4:) \cdots \exp(:f_{D+1}:)$ , the *nonlinear* part of  $\mathcal{M}$ , is more difficult to evaluate. Several approaches are currently available.

First, a Taylor series map  $\mathcal{T}$  of the form (2) in Sec.2.3.5, truncated beyond terms of degree  $D$ , may be used. Although truncation generally violates the symplectic condition, such an approach is usually adequate for single-pass systems such as linacs, beam lines, telescopes, and final focus systems. Symplecticity can be improved by retaining more terms in the expansion (7). This can be done for each of the factors in the product  $\mathcal{N}$ . Alternatively, we can combine the generators  $:f_3:, :f_4:, \dots$  into a single exponent using the BCH series, and then use a truncated version of (7) to evaluate the exponential of this single exponent. This method often proves satisfactory for a modest number of turns in a circulating system.

For long-term tracking it is necessary to maintain the symplectic condition exactly. For element-by-element tracking, which is slow, this may be done by using symplectic integrators [4, 5] (Sec.2.3.8). When a one-turn or few-lump map is employed, which is much faster if feasible, mixed-variable generating function or Cremona map approximations to these maps may be used.

For  $2n$ -dimensional phase space there is a  $2n(4n+1)$  parameter family of generating functions [4, 8]. Among them the best for tracking purposes appears to be that of Poincaré. Form the quantities  $\Sigma = (z^f + z^i)/2$  and  $\Delta = z^f - z^i$ . In the Poincaré generating function approach the action of  $\mathcal{N}$  is represented by the implicit (and

guaranteed symplectic) map

$$z^f = z^i + J\partial_{\Sigma}F \quad (29)$$

where  $F(\Sigma, \Delta)$  is a polynomial that can be calculated by formula from the polynomials  $f_3, f_4 \dots$ . Alternatively,  $J\partial_{\Sigma}F$  can be computed algorithmically from the associated Taylor map  $\mathcal{T}$ . The relations (29) can be solved numerically using Newton's method. There are three advantages to the use of the Poincaré generating function: First, suppose  $\mathcal{T}$  is initially inverted and then Poincaré symplectified. The result is the same as that found when  $\mathcal{T}$  is initially symplectified and the result then inverted. Second, suppose the map  $\mathcal{R}\mathcal{T}\mathcal{R}^{-1}$  is formed, where  $\mathcal{R}$  is *any* linear symplectic map, and then Poincaré symplectified. The result is the same as symplectifying  $\mathcal{T}$  and then sandwiching the result between  $\mathcal{R}$  and  $\mathcal{R}^{-1}$ . We may say that Poincaré symplectification is *covariant* under inversion and linear symplectic transformations. Finally, it can be shown that the Jacobian of  $\mathcal{T}$ , which is the identity matrix  $I$  for  $z^i = 0$ , can take on values all the way out to  $-I$  before the Poincaré symplectification process fails.

A Cremona map  $\mathcal{C}$  is a map that is both polynomial and exactly symplectic [9]. A Cremona map approximation consists of a series of Cremona maps  $\mathcal{C}_1, \mathcal{C}_2 \dots \mathcal{C}_m$  such that  $\mathcal{N} \simeq \mathcal{C}_1\mathcal{C}_2 \dots \mathcal{C}_m$ . The construction of satisfactory Cremona approximations requires a knowledge of group theory and cubature formulae on the manifolds  $S^2$  and  $SU(3)/SO(3)$ . In general the number of maps  $m$  depends on the degree  $D$ . For 4 and 5 dimensional phase space one has  $m = 12, 24, 30, 32, 50$ , and  $72$  for  $D = 4, 6, 7, 8, 10$ , and  $13$ , respectively. Because Cremona maps are polynomial and exactly symplectic, they can be evaluated rapidly on a computer to give results that are symplectic to machine precision. It has been found for the LHC that tracking using a Cremona approximation to the one-turn map satisfactorily reproduces its dynamic aperture, and is at least an order of magnitude faster than element-by-element tracking.

*Moments and emittance analysis:* Suppose  $h(z)$  is the initial distribution function describing a collection of particles in phase space [3, 10]. Define initial moments  $Z_{abc\dots}^i$  by the rule

$$Z_{abc\dots}^i = \langle z_a z_b z_c \dots \rangle = \int d^6z h(z) z_a z_b z_c \dots \quad (30)$$

Alternatively let  $P_{\alpha}(z)$ , where  $\alpha$  is some running index, denote a complete set of homogeneous

polynomials in  $z$  through terms of some fixed degree. Then one can define initial moments  $m_{\alpha}^i$  by the rule

$$m_{\alpha}^i = \int d^6z h(z) P_{\alpha}(z) \quad (31)$$

Next suppose the particle distribution is transported through some system described by a map  $\mathcal{M}$ . Then the final distribution at the end of the system is  $h(\mathcal{M}^{-1}z)$ . Correspondingly, the final moments are given by

$$\begin{aligned} m_{\alpha}^f &= \int d^6z h(\mathcal{M}^{-1}z) P_{\alpha}(z) \\ &= \int d^6z h(z) P_{\alpha}(\mathcal{M}z) \end{aligned} \quad (32)$$

Since the  $P_{\alpha}$  are complete, there is an expansion of the form

$$P_{\alpha}(\mathcal{M}z) = \sum_{\beta} \mathcal{D}_{\alpha\beta}(\mathcal{M}) P_{\beta}(z) \quad (33)$$

where the  $\mathcal{D}_{\alpha\beta}(\mathcal{M})$  are coefficients that can be calculated for any transfer map  $\mathcal{M}$ . It follows that moments transform linearly according to the rule

$$m_{\alpha}^f = \sum_{\beta} \mathcal{D}_{\alpha\beta}(\mathcal{M}) m_{\beta}^i \quad (34)$$

Note that by this method one can find the evolution of moments *without* tracking particle distributions. We also remark that with the aid of  $\mathcal{A}$  one can manufacture matched beam distributions, even in the nonlinear case, without tracking.

Let  $m$  be a vector with components  $m_{\alpha}$ , and let  $D(\mathcal{M})$  be a matrix with entries  $D_{\alpha\beta}(\mathcal{M})$ . Write (34) in the more compact form  $m^f = D(\mathcal{M})m^i$ . A function of moments  $I[m]$  is said to be a *kinematic invariant* if it obeys the relation  $I[D(\mathcal{M})m] = I[m]$  for all symplectic maps  $\mathcal{M}$ . Rather little is known about the existence and properties of kinematic invariants for the set of all symplectic maps. However, all kinematic invariants have been found when the symplectic maps  $\mathcal{M}$  are restricted to be linear [10]. Of particular interest are kinematic invariants that can be constructed from the second-order moments  $Z_{ab}^{[2]}$ . For a given particle distribution, Let  $Z^{[2]}$  be the  $6 \times 6$  matrix with entries  $Z_{ab}^{[2]}$ . For a 3-degree of freedom system, there are 3 such functionally independent invariants,

$$I_2^{(n)}[Z^{(2)}] = \text{tr}[(Z^{(2)}J)^n], \quad n = 2, 4, 6 \quad (35)$$

Even more can be said.  $Z^{[2]}$  can be shown to be positive definite, from which it follows

that there is a symplectic matrix  $A$  such that  $AZ^{[2]}A^T = \text{diag}(\lambda_1, \lambda_1, \lambda_2, \lambda_2, \lambda_3, \lambda_3)$  with all  $\lambda_j > 0$ . The quantities  $\lambda_j^2$  are mean-square *eigen* emittances that generalize the 1-D mean-square emittances  $\epsilon_i^2 = \langle q_i^2 \rangle \langle p_i^2 \rangle - \langle q_i p_i \rangle^2$  to the fully coupled case. In terms of the  $\lambda_j$  the kinematic invariants have the values

$$I_2^{(n)} = 2(-1)^{n/2}[\lambda_1^n + \lambda_2^n + \lambda_3^n], \quad n \text{ even} \quad (36)$$

There are symplectic matrix routines for finding  $A$  and the  $\lambda_j$ . If only the  $\lambda_j$  are required, they can be found from the eigenvalues of  $JZ^{[2]}$ . Finally, while the entries in  $Z^{[2]}$  evolve as a particle distribution propagates through a beam line, the eigen emittances remain *unchanged* (in the linear approximation). Thus, given an initial particle distribution, one can compute the initial second moments  $\langle z_a z_b \rangle^i$ , and from them the eigen emittances. Let  $\lambda_{\min}^2$  be the *minimum* of the quantities  $\lambda_j^2$ . Also, let  $\langle z_a z_b \rangle^f$  be the set of quadratic moments of the beam after it has passed through some beam processing system. Then there is a *classical uncertainty principle* and a *minimum emittance theorem* which state that, no matter what is done to the beam (ignoring nonlinear and nonsymplectic effects), there are the restrictions (for  $i = 1, 2, 3$ )

$$\langle q_i^2 \rangle^f \langle p_i^2 \rangle^f \geq \lambda_{\min}^2 \quad \text{and} \quad (\epsilon_i^2)^f \geq \lambda_{\min}^2 \quad (37)$$

This information is useful when designing sources (e.g. magnetic field immersed and/or otherwise tailored electron guns) and when designing a beam line to perform emittance manipulations on a beam. It should also be useful in analyzing the results of beam cooling experiments. In this case one can measure all quadratic moments before and after a cooling channel. Next compute the eigen emittances of  $Z^{[2]}$  before and  $Z^{[2]}$  after. Ideally, one would like to find that all the  $\lambda_j^2$  have decreased, or at least the minimum of the  $\lambda_j^2$  has decreased.

## References

- [1] A. Dragt, AIP Conf. Proc. 87, R. Carrigan et al ed. (1982)
- [2] A. Dragt et al, Ann. Rev. Nucl. Part. Sci. 38, 455 (1988)
- [3] A. Dragt et al, MaryLie 3.0 Users' Manual (2003); [www.physics.umd.edu/dsat/](http://www.physics.umd.edu/dsat/)
- [4] A. Dragt, Lie Methods for Nonlinear Dynamics with Applications to Accelerator Physics (2010); [www.physics.umd.edu/dsat/](http://www.physics.umd.edu/dsat/)
- [5] E. Forest, Beam Dynamics: A New Attitude and Framework, Harwood Academic (1998)
- [6] C. Mitchell, U. Maryland PhD Thesis (2007); [www.physics.umd.edu/dsat/](http://www.physics.umd.edu/dsat/)
- [7] C. Mitchell, A. Dragt, PRST-AB 13, 064001 (2010)
- [8] B. Erdélyi, Michigan State U. PhD Thesis (2001)
- [9] D. Abell, U. Maryland PhD Thesis (1995); [www.physics.umd.edu/dsat/](http://www.physics.umd.edu/dsat/); D. Abell, E. McIn-tosch, F. Schmidt, PRST-AB 6, 064001 (2003)
- [10] A. Dragt et al, PRA 45, 2572 (1992); N. Yampolsky et al, arXiv:1010.1558 (2010)

### 2.3.7 Differential Algebraic Techniques *M. Berz, Michigan State U.*

The study of the behavior of motion in the vicinity of a chosen reference solution is a central problem arising in many subfields of dynamical systems, including beam dynamics. The Taylor expansions of these solutions can be obtained by solving the so-called variational equations, which in beam physics has been carried out to orders two and three in the code Transport[1], to orders three for example in the codes TRIO [2], GIOS [3] and MaryLie [4], and to order five in COSY 5.0 [5]. This approach is laborious in practice, and the development of the DA techniques has greatly simplified this endeavor in beam physics and other fields. In their latest versions [6, 7, 8, 9], the unprecedented accuracy these methods afford for the solution of differential equations has been awarded the Moore prize for rigorous computing.

A review of DA as used in our field can be found in [10]. The DA techniques allow the convenient computation of high-order Taylor expansions of the transfer map  $\mathcal{M}$  which relates final particle coordinates  $\vec{z}_f$  to initial coordinates  $\vec{z}_i$  and parameters  $\vec{\delta}$ ,

$$\vec{z}_f = \mathcal{M}(\vec{z}_i, \vec{\delta}) \quad (1)$$

relative to one (in most beamlines, microscopes, or synchrotrons) or several (in FFAGs, cyclotrons and some spectrographs) reference orbits. The map  $\mathcal{M}$  is the *flow* (solution depending on initial conditions) of ODEs

$$\vec{z}' = \vec{f}(\vec{z}, s), \quad \text{independent variable } s \quad (2)$$

Since their introduction [11, 12], DA techniques have been utilized in most newly developed codes [13, 14, 15, 16, 17, 18] and the DA engines forming the core of COSY INFINITY [13] also form the backbone of various other codes. DA methods

have their origin in the algebraic study of problems involving differentiation and integration for the purpose of solving complicated integrals, ODEs, and PDEs. Pioneered by Liouville [19] in connection with the search of elementary integrals for elementary functions, it was put on a solid foundation and significantly enhanced by Ritt [20] and Kolchin [21]. Now the methods have gained prominence in the field of formula manipulation, where they provide the backbone of the theory of analytic quadrature and integration of ODEs [22].

For a given function  $f$  of  $v$  variables, we form a vector that contains all Taylor expansion coefficients at  $\vec{x} = \vec{0}$  up to a certain order  $n$ . Knowing this vector for two functions  $f$  and  $g$  allows to compute the respective vector for  $f + g$  and  $f \cdot g$ , since the derivatives of the sum and product function is uniquely defined from those of  $f$  and  $g$ . The resulting operations of addition and multiplication lead to an algebra, the Truncated Power Series Algebra (TPSA) [12, 23]. One can also introduce elementary functions like  $\exp$ ,  $\sin$  etc. on TPSA. TPSA allows the convenient computation of derivatives of any functional dependency on a computer [24].

For the solution of ODEs and PDEs it is necessary to introduce another operation. For any fixed function  $g$  with  $g(0) = 0$ , it is possible to determine the Taylor coefficients of  $g \cdot \partial f / \partial x_i = \partial_{g,i} f$  from those of  $f$ . Including the operation  $\partial_{g,i}$  and its inverse with addition and multiplication leads to a differential algebra (DA). Details on this particular DA can be found in [10, 24]. In passing it is worth mentioning that the DA structures also provide a novel way to a theory of differentials as infinitely small numbers [10]. For practical work with DA, care has to be taken to provide elementary operations  $+$ ,  $\cdot$ ,  $\partial_{g,i}$  that work to any order and any number of variables. Since usually many derivatives vanish due to symmetry, the active support of sparsity is essential [23].

**Solutions of ODEs and PDEs** To determine the map Eq.(1), it is necessary to solve the ODEs Eq.(2). Since usually fields are known only in the midplane, it is often also necessary to solve their PDEs in the process. The crudest approach to solve Eq.(2) is to replace all operations in a tracking code by the corresponding ones in DA [12, 11]. This replacement is similar in flavor to the modification of existing code through pre-processors performing “automatic differenti-

ation” [24], although in that field derivatives are typically only obtained to first or second order and the challenge lies in the efficient handling of large numbers of independent variables. An important practical problem of this map integration approach is to obtain higher order terms accurately, which can be done by reducing the step size. This approach has been used to retrofit several existing tracking codes, including recently MAD-X for the extraction of high-order DA maps.

However, using DA techniques it is possible to entirely avoid the approximate time stepping of conventional integrators, but rather develop new integrators of arbitrary order in time. Strikingly, they require only a single evaluation of the r.h.s per step, resulting in greatly increased efficiency and robustness [10, 6, 7]. One way is based on the common rewriting of the ODE as a fixed point problem

$$\vec{z}_f = \vec{z}_i + \int_{s_i}^{s_f} \vec{f}(\vec{z}, \tilde{s}) d\tilde{s} \quad (3)$$

Utilizing the operation  $\partial^{-1}$  for the integral, the problem can be iterated in DA with  $\tilde{s}$  as an additional variable. It can be shown that iteration converges to the exact result in  $n$  steps, where  $n$  is the order of the DA operations; moreover, this is not affected if in the  $i$ th iteration step the overall order is reduced to  $i$ . The result is an  $n$ th order integrator; for a given accuracy demand, the integrator typically results in a speed-up of about an order of magnitude. Moreover, the integrator also affords a rigorous and sharp estimate of all integration errors [6, 7, 8, 9].

Similarly, it is also possible to solve PDEs in finitely many steps. For this purpose, one eliminates differentiation with respect to one variable by integration. For example, the PDE

$$a_1 \frac{\partial V}{\partial x} (a_2 \frac{\partial V}{\partial x}) + b_1 \frac{\partial}{\partial y} (b_2 \frac{\partial V}{\partial y}) + c_1 \frac{\partial}{\partial z} (c_2 \frac{\partial V}{\partial z}) = 0 \quad (4)$$

which describes the scalar potential in particle optical relative coordinates, is rewritten as

$$V = V \Big|_{y=0} + \frac{1}{b_2} \int_y \left\{ \frac{\partial V}{\partial y} \Big|_{y=0} \right. \\ \left. - \int_y \frac{a_1}{b_1} \frac{\partial}{\partial x} (a_2 \frac{\partial V}{\partial x}) + \frac{c_1}{b_1} \frac{\partial}{\partial z} (c_2 \frac{\partial V}{\partial z}) \right\} \quad (5)$$

and again, iteration converges to the exact result in finitely many steps [10]. In addition to producing general PDE solvers, this approach allows the execution of the commonly performed out-of-plane expansion to arbitrary order [25].

It is important to observe that the accuracy of the out-of-plane expansion rests on the quality of field derivatives in the midplane. If the field arises from models, DA methods can directly be used to obtain these derivatives accurately to any order of interest. If on the other hand, the field is based on measurements, it is highly advantageous to represent the field in space through integrals over surface fields based on the Helmholtz theorem which resembles the Cauchy formula in complex analysis. DA methods can be used elegantly to expand fields simultaneously in the coordinates in the midplane, as well in the surface coordinates, and the integration is carried out directly utilizing the DA operation  $\partial^{-1}$ .

For ODEs that are time independent and for which  $\vec{z} = \vec{0}$  is a solution, which is the case in particle optical relative coordinates, another method can be applied. For a given function on phase space  $g(\vec{z}, s)$ , it is possible to obtain its derivative along the true solution via

$$\frac{d}{ds} g(\vec{z}, s) = \vec{f} \cdot \vec{\nabla} g + \frac{\partial}{\partial t} g = L_{\vec{f}} g \quad (6)$$

Apparently we also have  $d^n/ds^n g = L_{\vec{f}}^n g$ . If  $g$  is not explicitly time dependent, the operator  $L_{\vec{f}}^n$  can be evaluated directly within DA for any  $n$ ; using it for  $g = z_\nu$ , the components of the vector  $\vec{z}$ , we obtain an integrator of adjustable order. This method is utilized for the main fields of elements in COSY; an element is typically traversed in one step, and orders of 25-30 are usually chosen to obtain integration to nearly machine precision. Note that for certain particle optical systems, including the notoriously improperly treated fringe fields, there are other efficient and fast perturbative methods to obtain approximate solutions of high accuracy, like the method of symplectic scaling [26].

**Advanced DA operations and manipulation of maps** Given the  $n$ th order representations of two maps  $\mathcal{M}_n$  and  $\mathcal{N}_n$ , it is possible to determine the  $n$ th order representation of their composition  $\mathcal{M}_n \circ \mathcal{N}_n$  as long as  $\mathcal{N}_n$  has no constant parts. To this end, one simply inserts the components of  $\mathcal{N}_n$  into the polynomial defined by the coefficients of  $\mathcal{M}_n$ . This allows to determine the map of a combined system from the maps of the subsystems.

It is also possible to obtain the  $n$ th order representation  $\mathcal{N}_n$  of the inverse of  $\mathcal{M}_n$  as long as its linear part is invertible. To this end, one writes  $\mathcal{M}_n = \mathcal{M}_1 + \mathcal{M}_n^*$ , where  $\mathcal{M}_1$  is the linear part of

$\mathcal{M}_n$ . Then we have

$$\begin{aligned} \mathcal{I}_n &= (\mathcal{M}_1 + \mathcal{M}_n^*) \circ \mathcal{N}_n \\ &= \mathcal{M}_1 \circ \mathcal{N}_n + \mathcal{M}_n^* \circ \mathcal{N}_n \quad \Rightarrow \\ \mathcal{N}_n &= \mathcal{M}_1^{-1} (\mathcal{I} - \mathcal{M}_n^* \circ \mathcal{N}_n) \end{aligned} \quad (7)$$

a fixed point problem for  $\mathcal{N}_n$ . Beginning iteration with  $\mathcal{N}_n = \mathcal{I}_n$  yields convergence to the exact result in  $n$  steps because  $\mathcal{M}_n^*$  is purely nonlinear. The map of the reversion [27, 10] of a system, i.e. the system traversed backwards, is related to the inverse of the map; but since angles change directions when going backwards, the reversed map  $\mathcal{M}_r$  is  $\mathcal{M}_r = R \circ \mathcal{M}^{-1} \circ R$ , where  $R$  is the linear map that changes the signs of angles.

**Symplectic tracking with maps** Inverse maps can also be used to establish mixed-variable generating functions that represent the map, which are useful for symplectic tracking. Beginning from the  $n$ -th order map  $\mathcal{M}_n$ , we create  $\mathcal{N}_{n,1}$ , consisting of the position part of  $\mathcal{M}_n$ , and an identity for the momenta; thus  $(\vec{q}_f, \vec{p}_i) = \mathcal{N}_{n,1}(\vec{q}_i, \vec{p}_i)$ . Similarly we construct  $\mathcal{N}_{n,2}$  via  $(\vec{q}_i, \vec{p}_f) = \mathcal{N}_{n,2}(\vec{q}_i, \vec{p}_i)$ . Inversion of  $\mathcal{N}_{n,1}^{-1}$  (if possible) yields  $(\vec{q}_i, \vec{p}_i) = \mathcal{N}_{n,1}^{-1}(\vec{q}_f, \vec{p}_i)$ , and composition with  $\mathcal{N}_{n,2}$  from the right yields

$$(\vec{q}_i, \vec{p}_f) = \mathcal{N}_{n,2} \circ \mathcal{N}_{n,1}^{-1}(\vec{q}_f, \vec{p}_i) \quad (8)$$

From this mixed-variable representation, the underlying generating function can be obtained by a mere integration along a suitable path, again using the DA operation  $\partial^{-1}$ ; in a similar way as the case shown here, also all the other three common generating functions can be determined.

However, in practice it is frequently observed that different symplectification schemes result in different tracking results; so it is highly desirable to achieve symplectification with minimal modification of the prediction of Taylor transfer map. Utilizing Hofer's metric on spaces of Hamiltonians, this can actually be achieved in a unified way using DA techniques, resulting in the so-called EXPO symplectification scheme [28].

**Normal forms** Another important manipulation of maps is the transformation to normal form [10]. The first step is the transformation to the parameter-dependent fixed point  $\vec{z}(\vec{\delta})$  which satisfies  $(\vec{z}(\vec{\delta}), \vec{\delta}) = \mathcal{M}_n(\vec{z}(\vec{\delta}), \vec{\delta})$ . Subtracting the non-parameter identity map  $\mathcal{I}_n^{\vec{z}}$  on both sides we have  $(\vec{0}, \vec{\delta}) = (\mathcal{M}_n - \mathcal{I}_n^{\vec{z}})(\vec{z}(\vec{\delta}), \vec{\delta})$ , and thus

$$(\vec{z}(\vec{\delta}), \vec{\delta}) = (\mathcal{M}_n - \mathcal{I}_n^{\vec{z}})^{-1}(0, \vec{\delta}) \quad (9)$$

from which we read off  $\vec{z}(\vec{\delta})$  in the non-parameter lines.

The linear part of the normal form algorithm itself is based on a study of the eigenvalue spectrum of the map; if they are distinct and complex, then it amounts to a diagonalization to the form  $\mathcal{R}$ , where the phases of the diagonal terms correspond to the tunes. The nonlinear part of the normal form algorithm consists of a sequence of transformations  $\mathcal{S}_n$  that to order  $n$  have the form  $\mathcal{S}_n = \mathcal{I} + S_n$ . The orders higher than  $n$  in  $\mathcal{S}_n$  can in principle be picked freely, and frequently they are chosen so that all transformations  $\mathcal{S}_n$  belong to the same symmetry group as the original map.

Up to order  $n$ , we have  $\mathcal{S}_n^{-1} =_n \mathcal{I} - S_n$ ; letting  $\mathcal{N}_n = \mathcal{R} + \mathcal{O}_n$  denote the map that has been normalized to order  $n$ , we have

$$\begin{aligned} \mathcal{N}_n &= \mathcal{S}_n \circ \mathcal{N}_{n-1} \circ \mathcal{S}_n =_n (\mathcal{I} + S_n) \circ \mathcal{N}_{n-1} \circ (\mathcal{I} - S_n) \\ &=_n \mathcal{R} + \mathcal{O}_n + \{S_n \circ \mathcal{R} - \mathcal{R} \circ S_n\} \end{aligned} \quad (10)$$

Apparently, the  $S_n$  in the commutator  $\{S_n \circ \mathcal{R} - \mathcal{R} \circ S_n\}$  can now be chosen to remove  $n$ th order terms in  $\mathcal{O}_n$ . In the symplectic case, it turns out that the remaining terms are just those that describe motion on circles, with a frequency that depends on the radius and possible parameters  $\vec{\delta}$ , allowing direct computation of amplitude dependent tune shifts.

The radius vectors in each phase space pair represent invariants, the quality of which is directly determined by the magnitude of the coefficients of  $S_n$ , the so-called resonance strengths. Trying to minimize their size is an efficient way to enhance the overall nonlinear behavior of the system.

**Other DA-based tools** It is also possible to obtain Lie factorizations,

$$\begin{aligned} \mathcal{M}_n &=_n \{\mathcal{M}_1 \exp(: f_3 :) \exp(: f_4 :) \dots\} \mathcal{I} \\ \text{or } \mathcal{M}_n &=_n \{\dots \exp(: f_4 :) \exp(: f_3 :) \mathcal{M}_1\} \mathcal{I} \end{aligned} \quad (11)$$

of symplectic maps  $\mathcal{M}_n$  [10], where  $=_n$  means two functions are equal up to order  $n$ ; in fact, the DA approach currently represents the only method to obtain them to arbitrary order. The procedure is order-by-order; in the  $i$ th step, all orders less than  $i$  have already been taken care of, and the problem is reduced to finding  $f_{i+1}$  such that  $\exp(: f_{i+1} :) \mathcal{I} = \mathcal{I} + \vec{\nabla} f_{i+1} \cdot \hat{J}$  agrees with the given map to order  $i$ . This is an integration problem similar to the case of the generating functions above, which is resolved once more with the DA operator  $\partial^{-1}$ .

Some other DA-based methods not discussed above shall be mentioned briefly. For the analysis of spin dynamics, there are methods that allow the computation of the spin map, spin tracking, as well as invariant spin axis  $\hat{n}$ . There are also extensions of the DA methods to allow for a mathematically rigorous treatment of the remainder terms of Taylor's formula [6], which allow for the development of rigorous error bounds for integrators [7, 8] and also a rigorous bounding of stability times of nonlinear motion [29, 6] when combined with methods of verified global optimization [6]. There are methods for the solution of algebraically constrained differential equations, so-called differential algebraic equations [30], and methods to develop high order versions of the Fast Multipole Method for 3-D space charge computation [31].

**Applications** The differential algebraic methods form the core tools for the computation of maps and aberrations for a large number of codes. First developed and used in the code COSY INFINITY [13] which currently has about 1,500 registered users, the methods also form the core engines in many other codes [14, 15, 16, 17, 18].

The methods have been used for design, optimization and analysis of numerous beam optical devices, of which we only showcase a limited number for reasons of space. Maps are being used for the efficient symplectic tracking of dynamics in synchrotrons resulting in speed increases for the LHC [32] and the Tevatron; at the latter, they have also been used for purposes of linear decoupling [33] and a rigorous estimate of long-term stability [31]. Another important application is the tracking and correction based on more sophisticated models for fringe fields or other nonlinearities in large acceptance rings [34], symplectic tracking in light sources under consideration of wigglers and undulators, all the way to various applications for novel accelerator types like the FFAG.

The methods have also been used for the design and correction of fragment separators and particle spectrographs, including the MSU FRIB, S800 and A1900, those at TJNAF, Catania, RIKEN and OSAKA. In addition to computation and hardware correction, a particularly important application is the reconstructive correction of aberrations in high resolution spectrographs [35]. This technique uses combinations of the map and its inverse to eliminate aberrations

by measurement data in two planes, i.e. knowledge of  $(x_f, y_f, a_f, b_f)$ . Since this is uniquely determined as a function of  $(y_i, \delta_i, a_i, b_i)$ , utilizing map inversion techniques discussed above yields reaction energy and angles to precisions that are sometimes two orders of magnitude higher than without correction.

Finally, over the last years, use is also being made of the methods in electron microscopy, where they afford the ability to compute and correct high-order aberrations when the fields of the devices are known.

## References

- [1] K.L. Brown, SLAC Tech. Rep. 91 (1979)
- [2] T. Matsuo et al, Mass Spectrometry 24
- [3] H. Wollnik et al, AIP CP 177 (1988) 74
- [4] A.J. Dragt et al, IEEE Tran. Nuc. Sci. NS-3, 5, 2311 (1985)
- [5] M. Berz et al, NIM A258 (1987) 402
- [6] K. Makino, PhD thesis, Michigan State U., MSUCL-1093 (1998)
- [7] M. Berz, K. Makino, Reliable Computing 4 (4) (1998) 361
- [8] K. Makino, M. Berz, Int. J. Diff. Eq. & Appl. 10, 4 (2005) 353
- [9] K. Jackson, Proc. 5-th Int. Workshop on Taylor Models, in preparation
- [10] M. Berz, Modern Map Methods in Particle Beam Physics, Academic Press, San Diego, 1999; also available at <http://bt.pa.msu.edu/pub/>, together with more recent publications.
- [11] M. Berz, PA 24, 109 (1989)
- [12] M. Berz, NIM A258, 431 (1987)
- [13] M. Berz et al, <http://cosyinfinity.org>
- [14] Y. Yan, AIP Conf. 297 (1993) p.279
- [15] J. van Zeijts, AIP Conf. 297 (1993) p.285
- [16] L. Michelotti, Fermilab Tech. rep. (1990)
- [17] W. Davis et al, Proc. Workshop on Nonlinear Effects in Accelerators, 1993
- [18] F.C. Iselin, AIP Conf. 297 (1993) p.325
- [19] J.F. Ritt, Integration in Finite Terms - Liouville's Theory of Elementary Methods, Columbia U. Press, 1948
- [20] J.F. Ritt, Differential Algebra, APS, Washington, DC, 1950
- [21] E.R. Kolchin, Differential Algebraic Groups, Academic, NY, 1985
- [22] R.H. Risch, APS Bulletin 76, 605 (1970)
- [23] M. Berz, Forward algorithms for high orders and many variables, Automatic Differentiation of Algorithms: Theory, Implementation and Application SIAM
- [24] M. Berz et al, SIAM 1996
- [25] S. Manikonda, PhD thesis, Michigan State U., 2006
- [26] G. Hoffstaetter, M. Berz, PR E54, 4
- [27] W. Wan, M. Berz, PR E54, 3, 2870 (1996)
- [28] B. Erdélyi, M. Berz, PRL 87, 11, 114302 (2001)
- [29] K. Makino et al, ECMI Newsletter 39
- [30] J. Hoefkens, PhD thesis, Michigan State U. (2001)
- [31] H. Zhang, M. Berz, NIM submitted
- [32] M.L. Shashikant, M. Berz, B. Erdélyi, IOP CP 175, 299 (2002)
- [33] P.V. Snopok et al, NIM 558, 142 (2005)
- [34] M. Berz, B. Erdélyi, K. Makino, PRST-AB 3, 124001 (2000)
- [35] M. Berz et al, PR C 47, 2, 537 (1993)

### 2.3.8 Symplectic Integration Methods *H.Yoshida, Nat. Astron. Obs., Japan*

The exact time evolution (solution) of a Hamiltonian system

$$\frac{dq}{dt} = \frac{\partial H}{\partial p}, \quad \frac{dp}{dt} = -\frac{\partial H}{\partial q} \quad (1)$$

is symplectic [1, 2, 3]. This means that the mapping from  $(q(0), p(0))$  to  $(q(t), p(t))$  along the solution conserves the symplectic 2-form  $dp \wedge dq \equiv \sum_i dp_i \wedge dq_i$  exactly. This symplectic 2-form represents the sum of oriented areas of phase space element projected onto the plane of  $(p_i, q_i)$ , and for 1-D systems, "symplectic" is identical to "area-preserving" property. Traditional integration methods, such as the Euler method and the classical Runge-Kutta method, do not respect this fact and after a long term integration, fictitious damping or excitation occurs, making the result of integration unreliable. One needs integration methods which are designed to keep the symplectic property of the original Hamiltonian flow. These integration methods are in general called the symplectic integrators (symplectic integration methods). With use of symplectic integrators the error of energy does not grow monotonically. This comes from the existence of a conserved quantity which is close to the original Hamiltonian [3, 4].

#### 2.3.8.1 Methods of realization

There are explicit symplectic schemes and implicit schemes for general Hamiltonian systems.

**Implicit schemes for general Hamiltonian systems** The classical 4th order Runge-Kutta

method is generalized as follows. For a given system of ordinary differential equations,

$$\frac{dz}{dt} = f(z) \quad (2)$$

coefficients  $k_i$  are determined by the simultaneous algebraic equations

$$k_i = f \left( z + \tau \sum_{j=1}^s a_{ij} k_j \right) \quad (3)$$

( $i = 1, \dots, s$ ). Then the one-step mapping  $z \rightarrow z'$  is defined by ( $\tau = \Delta t$ )

$$z' = z + \tau \sum_{j=1}^s b_j k_j \quad (4)$$

Here,  $a_{ij}$  and  $b_j$  are scalar constants which characterize the scheme. This scheme is called the Implicit Runge-Kutta (IRK) method in general.

If the given system is a Hamiltonian system (1), and if the constants  $a_{ij}$  and  $b_j$  satisfy the conditions

$$M_{ij} \equiv b_i a_{ij} + b_j a_{ji} - b_i b_j = 0, \quad (1 \leq i, j \leq s) \quad (5)$$

identically, then the mapping (4) is known to be symplectic [5, 2]. An implicit symplectic integration method is thus obtained.

The simplest case ( $s = 1$ ) which satisfies the conditions (5) is given by  $a_{11} = 1/2$ ,  $b_1 = 1$ , and we have the scheme,

$$k_1 = f \left( z + \frac{\tau}{2} k_1 \right), \quad z' = z + \tau k_1$$

or, more concisely,

$$z' = z + \tau f \left( \frac{z + z'}{2} \right) \quad (6)$$

which is known as the implicit midpoint rule. This scheme has 2nd order accuracy. For  $s = 2$ , an example is given by

$$(a_{ij}) = \begin{bmatrix} \frac{1}{4} & \frac{1}{4} - \frac{\sqrt{3}}{6} \\ \frac{1}{4} + \frac{\sqrt{3}}{6} & \frac{1}{4} \end{bmatrix}, \quad (b_j) = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \end{bmatrix} \quad (7)$$

and this scheme has order 4 accuracy.

For general Hamiltonian systems, it is also possible to design symplectic schemes using the “generating function” [6].

**Explicit schemes for special Hamiltonian systems** For Hamiltonians of the form

$$H = T(p) + V(q) \quad (8)$$

there exist explicit symplectic algorithms. The idea to construct explicit schemes is simply to approximate the original Hamiltonian flow by a composition of trivial symplectic mappings,

$$q' = q + c_i \tau \left( \frac{\partial T}{\partial p} \right), \quad p' = p \quad (9)$$

and

$$q' = q, \quad p' = p - d_i \tau \left( \frac{\partial V}{\partial q} \right) \quad (10)$$

repeatedly. Here constants  $(c_i, d_i)$ , ( $i = 1, \dots, k$ ) are determined so that the composed mapping  $(q, p) \rightarrow (q', p')$  coincides with the Taylor series expansion of the solution up to the desired order.

The simplest, 1st order scheme  $(q, p) \rightarrow (q', p')$  is given by

$$q' = q + \tau \left( \frac{\partial T}{\partial p} \right)_{p=p}, \quad p' = p - \tau \left( \frac{\partial V}{\partial q} \right)_{q=q'} \quad (11)$$

The 2nd order scheme  $(q, p) \rightarrow (q', p')$  is attained by

$$q^* = q + \frac{\tau}{2} \left( \frac{\partial T}{\partial p} \right)_{p=p}, \quad p' = p - \tau \left( \frac{\partial V}{\partial q} \right)_{q=q^*},$$

$$q' = q^* + \frac{\tau}{2} \left( \frac{\partial T}{\partial p} \right)_{p=p'} \quad (12)$$

and this scheme is known as the leap-frog method.

This problem to derive explicit symplectic integrators is well reformulated in terms of Lie algebraic language (Sec.2.3.6). First rewrite the Hamilton equation (1) in the form

$$\frac{dz}{dt} = [z, H(z)] \quad (13)$$

where brackets stand for the Poisson bracket. If we introduce a differential operator  $:G:$  by  $:G:F \equiv [G, F]$ , then (13) is written as  $\dot{z} = -:H:z$ , so the formal exact solution of  $z(t)$  from  $t = 0$  to  $t = \tau$  is given by

$$z(\tau) = [\exp(-\tau:H:)]z(0) \quad (14)$$

For a Hamiltonian of the form (8), we have

$$z(\tau) = \exp[\tau(A + B)]z(0) \quad (15)$$

where  $A \equiv -:T:$  and  $B \equiv -:V:.$  Operators  $A$  and  $B$  do not commute in general.

Suppose  $(c_i, d_i)$  is a set of constants which satisfies  $\exp[\tau(A + B)] =$

$$\prod_{i=1}^k \exp(c_i \tau A) \exp(d_i \tau B) + \mathcal{O}(\tau^{n+1}) \quad (16)$$

for a given integer  $n$ , which corresponds to the order of the integrator. Then consider a mapping from  $z = z(0)$  to  $z' = z(\tau)$ , given by

$$z' = \left[ \prod_{i=1}^k \exp(c_i \tau A) \exp(d_i \tau B) \right] z \quad (17)$$

This mapping is symplectic because it is a product of elementary symplectic mappings, and approximates the exact solution (15) up to the order of  $\tau^n$ . Furthermore (17) is explicitly computable. In fact (17) gives the succession of the mappings

$$\begin{aligned} q_i &= q_{i-1} + \tau c_i \left( \frac{\partial T}{\partial p} \right)_{p=p_{i-1}} \\ p_i &= p_{i-1} - \tau d_i \left( \frac{\partial V}{\partial q} \right)_{q=q_i} \end{aligned} \quad (18)$$

for  $i = 1$  to  $i = k$ , with  $(q_0, p_0) = z$  and  $(q_k, p_k) = z'$ . An  $n$ -th order symplectic integrator is thus realized. For example, when  $n = 1$ , a trivial solution is  $c_1 = d_1 = 1$ , ( $k = 1$ ), which corresponds to the identity

$$\exp[\tau(A + B)] = \exp(\tau A) \exp(\tau B) + \mathcal{O}(\tau^2) \quad (19)$$

and gives the 1st order integrator (11). When  $n = 2$ , we find a solution  $c_1 = c_2 = 1/2$ ,  $d_1 = 1$ ,  $d_2 = 0$ , ( $k = 2$ ), i.e.  $\exp[\tau(A + B)] =$

$$\exp\left(\frac{\tau}{2}A\right) \exp(\tau B) \exp\left(\frac{\tau}{2}A\right) + \mathcal{O}(\tau^3) \quad (20)$$

and yields the 2nd order integrator (12).

**Composition method to obtain higher order schemes** Once a lower order scheme is given, a higher order scheme is obtained by a composition of lower order ones [7, 3].

For example, let  $S_2(\tau)$  be a 2nd order symplectic scheme, (12) or (6). Then a 4th order integrator  $S_4(\tau)$  is obtained by a symmetrical composition of 2nd order ones as

$$S_4(\tau) = S_2(x_1 \tau) \circ S_2(x_0 \tau) \circ S_2(x_1 \tau) \quad (21)$$

with

$$x_0 = \frac{-2^{1/3}}{2 - 2^{1/3}}, \quad x_1 = \frac{1}{2 - 2^{1/3}} \quad (22)$$

In order to have a 6th order integrator,  $S_6(\tau)$ , composition of 7 2nd order one is necessary:

$$\begin{aligned} S_6(\tau) &= S_2(y_3 \tau) \circ S_2(y_2 \tau) \circ S_2(y_1 \tau) \circ S_2(y_0 \tau) \\ &\circ S_2(y_1 \tau) \circ S_2(y_2 \tau) \circ S_2(y_3 \tau) \end{aligned} \quad (23)$$

with numerical constants

$$\begin{aligned} y_1 &= -1.1776 \ 79984 \ 17887 \\ y_2 &= 0.23557 \ 32133 \ 59357 \\ y_3 &= 0.78451 \ 36104 \ 77560 \\ y_0 &= 1 - 2(y_1 + y_2 + y_3) \end{aligned} \quad (24)$$

Much higher order integrator can be obtained by a composition of lower order ones in a similar way.

### 2.3.8.2 Symplectic method vs. nonsymplectic method

For the harmonic oscillator,  $H = \frac{1}{2}(p^2 + q^2)$ , the Euler method (nonsymplectic) gives the mapping

$$\begin{bmatrix} q' \\ p' \end{bmatrix} = \begin{bmatrix} 1 & \tau \\ -\tau & 1 \end{bmatrix} \begin{bmatrix} q \\ p \end{bmatrix} \quad (25)$$

and one finds easily that at each step, the value of energy is multiplied by  $(1 + \tau^2)$ , i.e.,  $(p'^2 + q'^2) = (1 + \tau^2)(p^2 + q^2)$ . On the other hand, the 1st order symplectic scheme (11) gives the mapping

$$\begin{bmatrix} q' \\ p' \end{bmatrix} = \begin{bmatrix} 1 & \tau \\ -\tau & 1 - \tau^2 \end{bmatrix} \begin{bmatrix} q \\ p \end{bmatrix} \quad (26)$$

and the energy does not grow monotonically. This is because (26) has the exact conserved quantity

$$\tilde{H} = H + \tau H_1 = \frac{1}{2}(p^2 + q^2) + \frac{\tau}{2}pq \quad (27)$$

and the numerical solution cannot deviate from the true solution curve. For general Hamiltonians, there exists always a conserved quantity

$$\tilde{H} = H + \tau H_1 + \tau^2 H_2 + \tau^3 H_3 + \dots \quad (28)$$

which is given as a formal power series of  $\tau$ , as far as symplectic algorithm is employed. This prevents the occurrence of fictitious damping or excitation, which is common when a nonsymplectic method is used.

## References

- [1] E. Hairer, S.P. Nørsett, G. Wanner, Solving Ordinary Differential Equations (2nd rev. ed.), Springer (1993)
- [2] J.M. Sanz-Serna, M.P. Calvo, Numerical Hamiltonian Problems, Chapman & Hall (1994)
- [3] H. Yoshida, Celest. Mech. 56 (1993) 27
- [4] G. Benettin, A. Giorgilli, J. Statistical Phys. 74 (1994) 1117
- [5] J.M. Sanz-Serna, BIT 28 (1988) 877
- [6] P.J. Channell, J.C. Scovel, Nonlinearity 3 (1990) 231
- [7] H. Yoshida, PL A150 (1990) 262

### 2.3.9 Dynamic Aperture

A. Wolski, U. Liverpool

As a particle travels around a storage ring, the amplitude of its motion with respect to the closed orbit may increase due to certain dynamical effects. The dynamic aperture may be defined in terms of the variation in amplitude over time: a particle is

said to be outside the dynamic aperture of the storage ring if its amplitude exceeds some large value after a sufficient number of turns.

Although it usually refers to an idealized or simplified system, dynamic aperture is a useful concept because it characterizes dynamical effects that may limit the acceptance of a storage ring. The acceptance refers to the range of trajectories of all particles that can remain indefinitely in a storage ring, taking into account physical limitations (e.g. vacuum chamber aperture) as well as dynamical effects, and is important for injection efficiency and beam lifetime.

Although analytic formulae have been developed for some cases [1, 2], dynamic aperture is generally computed using long-term (many hundreds, or thousands of turns) tracking. Usually, only single-particle dynamical effects are considered, but collective effects (e.g. beam-beam forces, Sec.2.5.1) may be included. Longitudinal dynamics may be included in different ways: they may be ignored altogether; it may be assumed that the energy deviation is constant; or, an accurate model including synchrotron motion may be used. For lepton storage rings, synchrotron radiation (Sec.3.1) can have a significant impact on the dynamic aperture, because of damping and quantum excitation. However, including radiation effects in the model can significantly slow the tracking, so such effects are often neglected or checked only once at the end.

The main issues associated with dynamic aperture are as follows. First, the computational results can have a strong dependence on the model (which dynamical effects are included, what boundaries are used to determine unstable trajectories, number of turns tracked, etc.) Second, long-term tracking can be computationally expensive. Third, it is difficult to represent the dynamic aperture (which, in its most general sense, refers to a set of points in 6-D phase space) in a complete and meaningful way. Finally, although some techniques have been developed for identifying and minimizing limitations on dynamic aperture, the problem of optimizing the dynamic aperture subject to other design constraints has not been completely solved.

**Factors affecting dynamic aperture** A storage ring lattice in which the equations of motion are perfectly linear and lead to stable oscillations has infinite dynamic aperture. The introduction of nonlinear elements will generally have two

effects: first, the elements will cause oscillations about the closed orbit to grow in amplitude for particular values of the tunes (resonances); second, the tunes will depend on the amplitudes of the oscillations. The result will be a limited dynamic aperture. In addition, ripple in magnet strengths due to power supply defects will affect the dynamic aperture through the existing nonlinearities.

Nonlinearities can come from a variety of sources. First, the equations of motion for particles in “linear” elements such as dipoles and quadrupoles are only linear under certain approximations. Second, nonlinear elements such as sextupoles may be deliberately introduced to correct such effects as chromaticity (chromatic sextupoles are often the dominant limitation on dynamic aperture). Third, all magnets in practice have fields with higher-order components. These components may be intrinsic in the design (systematic errors), or the result of variations in the construction (random errors).

The behaviour of nonlinear systems can be very complicated, but some understanding in the case that the nonlinearity is a perturbation applied to an integrable system is provided by the Kolmogorov-Arnol'd-Moser (KAM) theorem [3, 4]. The KAM theorem states that for a weak perturbation, periodic trajectories sufficiently far from resonant frequencies will remain stable (albeit with modified frequencies) while other trajectories close to resonance become unstable (Arnol'd diffusion [4]). The Chirikov criterion [5] states that where resonances overlap, particle motion becomes chaotic.

**Computing dynamic aperture** In principle, any tracking code can be used to compute the dynamic aperture of a storage ring. Tracking codes in common use include: MAD [6], DIMAD [7], BMAD [8], SIXTRACK [9], LEGO [10], Merlin [11], and AT [12]. The first four of these programs follow instructions prepared in a command file. LEGO and Merlin are C++ class libraries. AT is a set of tracking and beamline analysis routines working within Matlab.

A basic estimate of dynamic aperture can be obtained in two steps. First, the lattice is defined with the required configuration (magnet strengths, rf voltage, magnetic field and alignment errors etc.) Second, particles are tracked over a range of initial conditions. The initial conditions of those particles whose trajectories stay within specified

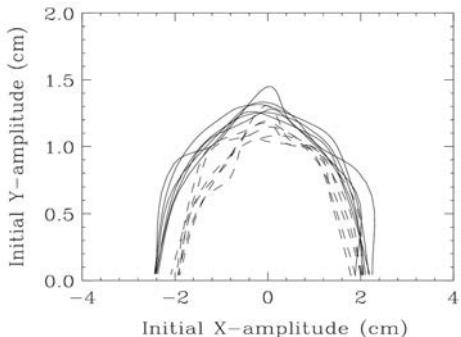


Figure 1: Dynamic aperture computation for SPEAR3 [13], for six seeds of machine errors. Solid lines: zero energy deviation. Dashed line: 3% energy deviation.

limits (usually taken at the limit of validity for the maps used for individual components) for a specified number of turns are deemed to lie within the dynamic aperture.

Performing the tracking to carry out a thorough exploration of phase space can be very time consuming, and can be a significant issue if many lattice configurations (and sets of random errors) need to be studied. The computation time can be reduced either by reducing the size of the problem, or by using techniques to speed the tracking.

The most common way to reduce the size of the problem is to limit the volume, or the dimensionality (or both) of phase space used for initial conditions. For example, a search can be made for the largest initial amplitudes in transverse coordinate space, for which the subsequent trajectories are stable (Fig.1). However, there is a risk that the search procedure may miss points within the final boundary that lead to unstable trajectories: note that the dynamic aperture is not necessarily a connected region in phase space.

In the case that a storage ring lattice includes only normal (i.e. not skew) multipole magnets, then the equations of motion are symmetric under the transformation  $y \mapsto -y$  (though not under  $x \mapsto -x$ ), where  $y$  is the vertical coordinate. In that case, tracking is necessary to cover only  $y > 0$ .

If damping processes are present (e.g. synchrotron radiation, or adiabatic damping during acceleration), then there is little value in tracking for more than one or two damping times. However, in hadron storage rings operating at fixed energy, it may be necessary to track for tens or

hundreds of thousands of turns. It is also possible to estimate the dynamic aperture from a *survival plot*, showing the number of turns a particle survives, as a function of the initial (coordinate) amplitude: the data points can often be fitted by a line asymptotic to some amplitude, below which particles remain in the storage ring indefinitely [14]. For long-term tracking, accuracy is important, since small errors can become significant over many turns.

Since computation of the dynamic aperture usually includes only single-particle effects, parallelization can reduce the computation time in proportion to the number of processors available. Another way to speed the computation is to represent groups of elements (or even the entire lattice) by a single dynamical map [15]. However, computation of the map can be time-consuming, since it is often necessary to include terms to very high order to achieve sufficient accuracy. Unless the maps are symplectic (i.e. the symplectic error is small, compared to physical nonsymplectic effects such as radiation) then artificial growth or damping of oscillation amplitudes can lead to a poor prediction of the dynamic aperture. To enforce symplecticity, it is possible to use symplectic representations of dynamical maps, such as Lie transforms, or mixed-variable generating functions (Secs.2.3.6, 2.3.8); however, this adds complexity, and such maps can be difficult or expensive to compute (and evaluate) to sufficiently high order.

**Measuring dynamic aperture** An experimental estimate of the dynamic aperture in an operating storage ring may be obtained by observing the particle loss that results from kicking a bunch to excite coherent oscillations at different amplitudes. It can be difficult to distinguish dynamical effects from the effects of physical apertures in such experiments, although use of adjustable collimators can give information. Studies (e.g. at ESRF [16]) have shown that there can be good agreement between the measured and computed dynamic apertures, if a sufficiently complete and detailed model is used for the computation. See also Sec.2.3.12.

**Characterizing the underlying dynamics** During the design process, it is often important to understand the reasons for the dynamic aperture limits. This involves characterizing the underlying dynamics at a more detailed level, in particular to understand the tune shifts with

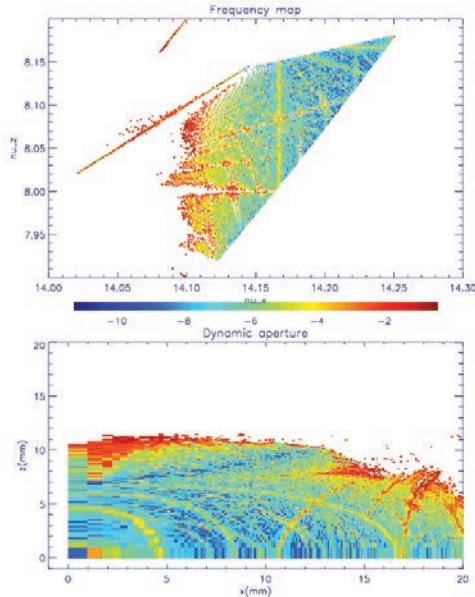


Figure 2: Frequency Map Analysis for the Advanced Light Source [22].

amplitude, and the strengths of resonances. A phase space portrait (showing the coordinates and momenta of a set of particles on successive turns) can give a simple indication of these features. Nonlinear effects appear as distortions of the set of ellipses that would be produced in a linear lattice; resonances may appear as “islands”, though this depends on the details of the dynamical map. Phase space portraits produce mostly qualitative information, and they may be difficult to interpret in the presence of coupling between different degrees of freedom.

The limits on the dynamic aperture may also be investigated by inspecting the strengths of various nonlinear terms in the dynamical map for a complete turn. The map may be computed, e.g., using a differential algebra code (e.g. COSY [17], Sec.2.3.7) or a Lie algebra code (e.g. Zlib [18], or MARYLIE [19], Sec.2.3.6). The strengths of tune shifts and resonance driving terms may be deduced from different terms in the map.

One of the most powerful and widely-used techniques for characterizing the dynamics underlying the dynamic aperture is *frequency map analysis* (FMA [20]). FMA involves tracking a single particle over many turns, starting with specified

values of the coordinates, and then determining the frequency of oscillation around the closed orbit with high precision. Strictly speaking, the tunes are only defined for regular orbits, lying on an invariant torus (an  $N$ -dimensional closed surface in  $2N$ -dimensional phase space). However, numerical techniques can be used to determine oscillation frequencies even for quasi-periodic motion. For example, using numerical analysis of the fundamental frequencies (NAFF [21]), one searches for the quantities  $\nu$  and  $\phi_0$  that maximize the sum  $\sum_{n=1}^N x_n \cos(2\pi\nu \frac{n}{N} + \phi_0)$ , where  $x_n$  is the coordinate on turn  $n$ . Applying NAFF to a range of initial coordinates, and plotting the horizontal and vertical frequencies on a tune grid produces a frequency map, Fig.2. Tune shifts with amplitude are visible from the area of tune space covered. Resonance strengths can be seen from the clustering of points around resonance lines. The “diffusion rate” in tune space (the rate of change of the tunes over the period of the tracking) can be indicated on a color scale. Regions of poor stability (where small errors are likely to lead to trajectories becoming unstable) are characterized by high diffusion rates, and are often associated with the intersection of resonance lines in tune space. One of the strengths of FMA lies in the indications it gives for ways to improve the dynamic aperture; for example, it may be possible to avoid a strong resonance simply by making a change in the working point in tune space (though such a change may also cause changes in the tune shifts with amplitude that lead to particles at large amplitudes crossing other resonances).

Dynamical stability can also be characterized by calculating the *Lyapunov exponent*, which describes the rate at which two initially close trajectories diverge from each other [14].

**Optimizing dynamic aperture** A design goal for any storage ring is to achieve sufficient acceptance to meet injection efficiency and lifetime criteria. The dynamic aperture must be large enough to meet the required acceptance, with some margin provided e.g. in case magnet errors are larger than expected. Despite various attempts to develop systematic procedures for improving the dynamic aperture subject to various other design constraints, a definitive technique has not yet been produced. It is usually necessary to approach each lattice on an individual basis, although there are some general rules and principles that can be applied.

One key aspect is the choice of working point: this can be selected, for example, based on the results of FMA. Alternatively, a tune scan can be performed, with the size of the dynamic aperture at different working points represented by points of different sizes or colors on a tune diagram (a *swamp plot*).

If the tune shifts with amplitude are large and there are strong resonances, it may not be possible to achieve sufficient dynamic aperture at any working point. Then, it is necessary to reduce the combined strengths of the nonlinear components in the lattice. The need for strong chromatic sextupoles can be avoided by designing the linear lattice for low chromaticity (small beta functions and quadrupole strengths), and by locating the sextupoles at appropriate points (high dispersion, good separation between the beta functions) to maximize their chromatic effects. In some cases, it may be possible to design the chromatic correction so that sextupoles occur in pairs with  $\pi$  phase advance (in  $x$  and  $y$ ) between them (a  $-I$  transformer) [23]; in that case, for particles with zero energy deviation, the geometric aberrations (tune shifts with amplitude, resonance terms) cancel out.

Chromaticity can make it particularly difficult to achieve good dynamic aperture in low-emittance storage rings (e.g. for third-generation synchrotron light sources, Sec.2.2.4) or colliders. Such machines generally require low dispersion and low beta functions in the dipoles; if the lattice also needs to be compact, then it can be difficult to provide good locations for sextupoles for chromatic correction. Colliders generally need very low beta functions at the interaction points, and the beta functions can reach very large values in the final quadrupoles, resulting in high chromaticity.

It may be beneficial to include nonlinear components in the lattice (e.g. sextupoles at locations with low or zero dispersion; octupoles etc.) specifically to provide control over the nonlinear dynamics. However, reducing certain nonlinear terms in the map often comes at the cost of introducing additional higher-order terms, that may lead to an overall reduction in dynamic aperture.

## References

- [1] G. Guignard, J. Hagel, Lecture Notes in Physics 247, 367, Springer (1986)
- [2] J. Gao, NIM A451 (2000) 545
- [3] V.I. Arnol'd, Mathematical Methods of Classical Mechanics, Springer (1978)
- [4] M. Tabor, Chaos and Integrability in Nonlinear Dynamics: An Introduction, Wiley (1989)
- [5] B.V. Chirikov, J. Nucl. Energy, Part C Plasma Phys. 1, 253 (1960)
- [6] <http://mad.web.cern.ch/mad/>
- [7] <http://www.slac.stanford.edu/accel/ilc/codes/dimad/>
- [8] <http://www.lepp.cornell.edu/dcs/bmad/>
- [9] F. Schmidt, CERN SL/94-56 (AP), 1994
- [10] Y. Cai et al, PAC 1997, 2583
- [11] <http://www.desy.de/merlin/>
- [12] <http://www-ssrl.slac.stanford.edu/at/>
- [13] J. Corbett et al, PAC 1999, 2364
- [14] M. Giovannozzi, W. Scandale, E. Todesco, AIP Proc. 395 (1997) 243
- [15] A.W. Chao et al, PAC 1991, 336
- [16] A. Ropert, PAC 1998, 921
- [17] [http://bt.pa.msu.edu/index\\_cosy.htm](http://bt.pa.msu.edu/index_cosy.htm)
- [18] Y.T. Yan, PAC 1991, 333
- [19] <http://www.physics.umd.edu/dsat/dsatmarylie.html>
- [20] D. Robin et al, PRL 85 (2000) 558
- [21] J. Laskar, Icarus 88 (1990) 266
- [22] C. Steier et al, EPAC 2000, 1077
- [23] Y. Cai et al, PAC 1995, 576

### 2.3.10 Decoherence

*M.A. Furman, LBNL*

If a stored beam is kicked transversely by angle  $\Delta x'$  (or is injected offset), its centroid betatron signal decoheres due to betatron tune spread. We define  $q \equiv x/\sigma_x$  and  $p \equiv (\alpha_x x + \beta_x x')/\sigma_x$  where  $\alpha_x$ ,  $\beta_x$  and  $\sigma_x$  are the lattice functions and rms beamsize at the observation point (=kick-point). We assume: (1) the beam is Gaussian in  $(x, x')$  and in the relative momentum deviation  $\delta \equiv \Delta P/P_0$ ; (2) there is no  $x$ - $y$  coupling and no synchro-betatron coupling; (3) damping, quantum excitation and the mutual interactions of the particles can be ignored; and (4) the tune dependence on amplitude and energy offset is

$$\nu = \nu_0 - \mu(q^2 + p^2) + \xi\delta \quad (1)$$

where  $\xi$  = chromaticity. Then the time evolution of the beam centroid is [1, 2]

$$\begin{aligned} \langle q \rangle + i \langle p \rangle &= \frac{i Z F(n)}{(1 - i\theta)^2} \\ &\times \exp \left( -2\pi i n \nu_0 + \frac{Z^2}{2} \frac{i\theta}{1 - i\theta} \right) \end{aligned} \quad (2)$$

where  $n$  = turn number,  $\theta = 4\pi\mu n$ ,  $Z = \beta_x \Delta x' / \sigma_x$  and the chromatic form factor

$$F(n) = \exp \left[ -2 \left( \frac{\xi \sigma_\delta}{\nu_s} \right)^2 \sin^2(\pi n \nu_s) \right] \quad (3)$$

where  $\nu_s$  = synchrotron tune. The second moments after the kick are

$$\begin{pmatrix} \langle q^2 \rangle \\ \langle qp \rangle \\ \langle p^2 \rangle \end{pmatrix} = \left( 1 + \frac{Z^2}{2} \right) \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} + \frac{Z^2 F^2(n)}{2(1+4\theta^2)^{3/2}} \exp \left( -\frac{2Z^2 \theta^2}{1+4\theta^2} \right) \begin{pmatrix} -\cos \psi \\ \sin \psi \\ \cos \psi \end{pmatrix} \quad (4)$$

$$\psi \equiv 4\pi n \nu_0 - \frac{Z^2 \theta}{1+4\theta^2} - 3 \tan^{-1} 2\theta \quad (5)$$

The normalized rms size is  $\sigma_x(n)/\sigma_x(0) = (\langle q^2 \rangle - \langle q \rangle^2)^{1/2}$ . Note that  $\langle q^2 \rangle + \langle p^2 \rangle = 2 + Z^2 = \text{constant}$ . The amplitude  $A \equiv |\langle q \rangle + i \langle p \rangle|$  of the beam centroid is

$$A(n) = \frac{ZF(n)}{1+\theta^2} \exp \left( -\frac{Z^2 \theta^2}{2(1+\theta^2)} \right) \quad (6)$$

Long after the kick,  $\theta \gg 1$ , the centroid amplitude decoheres as  $A \sim \theta^{-2}$ , while the rms beam size approaches an equilibrium  $\sigma_x(\infty)/\sigma_x(0) = (1 + Z^2/2)^{1/2}$ .

As time elapses,  $F(n)$  periodically comes back to its peak value of unity. Therefore, if  $\mu = 0$  (i.e.  $\theta = 0$ ), the beam centroid “recoheres” with the synchrotron period. This effect provides a way [3] to measure the product  $\xi \sigma_\delta$  (assuming  $\nu_s \ll 1$ ). If  $\mu \neq 0$ , the recoherence is still partially present.

The formulae above apply to 1-D. Extension to 2-D, including  $x$ - $y$  coupling in the tune dependence with amplitudes, is addressed in [4]. Ref.[5] treats the decoherence phenomenon including synchrobetatron coupling, damping and quantum excitation. Ref.[6] applies the canonical Hamiltonian perturbation formalism to 2-D decoherence in the presence of an arbitrary non-linear tune dependence on amplitudes; this formalism allows computing the decoherence rate of a beam trapped in a resonant island. Ref.[2] presents data on the dependence of decoherence rate on beam intensity in the SLC (in particular, the dependence on the sign of  $\xi$  through head-tail damping), while Ref.[7] analyzes this effect using a two-particle model. Experimental observation

of head-tail damping at the TRISTAN MR is analyzed in [8]. A full 3-D analysis is provided in [9], and was used as a tool to measure the emittance in the TRISTAN ring. An application to the Photon Factory is presented in [10], and to a version of the muon collider in [11]. A more general discussion is provided in [12].

## References

- [1] R.E. Meller et al, SSC-N-360, 1987
- [2] M.G. Minty et al, PAC95, p.3037
- [3] I.C. Hsu, PA 34, 43 (1990)
- [4] S.Y. Lee, Proc. Int. Workshop on Nonlinear Problems in Acc. Phys., Berlin, 1992, Inst. Phys. Conf. Series, p.249
- [5] H. Moshammer, PR E48(3), 2140 (1993)
- [6] J. Shi, S. Ohnuma, PAC 93, p.3603
- [7] G.V. Stupakov, A.W. Chao, PAC 95, p.3288
- [8] N. Akasaka, S. Kamada, EPAC 96, p.1141
- [9] S. Kamada, N. Akasaka, K. Ohmi, PAC 97, p.1831
- [10] Y. Kobayashi, K. Ohmi, EPAC 98
- [11] E.-S. Kim, M. Yoon, Jpn. J. Appl. Phys. 40, 4237 (2001)
- [12] K.-Y. Ng, Physics of Intensity-Dependent Beam Instabilities, World Scientific, 2006, Ch.13

### 2.3.11 Momentum Compaction and Phase Slip Factor

K.Y. Ng, FNAL

The phase slip factor  $\eta$  is the relative slip in revolution period  $T$  for a particle with fractional off-momentum  $\delta = \Delta p/p_0$ , i.e.  $\Delta T/T_0 = \eta\delta$ , where the subscript zero stands for on-momentum (Sec.2.1.2). The various orders of momentum-compaction factor  $\alpha_i$  give the relative increase in closed-orbit length  $C$  for an off-momentum particle, or  $\Delta C/C_0 = \sum_{i=0}^{\infty} \alpha_i \delta^{i+1}$ . With  $\eta = \sum_{i=0}^{\infty} \eta_i \delta^i$ , we have [1]

$$\eta_i = \alpha_i - \frac{\eta_{i-1}}{\gamma_0^2} + \frac{3\beta_0^2 \eta_{i-2}}{2\gamma_0^2} + \frac{(1-5\beta_0^2)\beta_0^2 \eta_{i-3}}{2\gamma_0^2} - \frac{5(3-7\beta_0^2)\beta_0^4 \eta_{i-4}}{8\gamma_0^2} + \dots, \quad \eta_i = \begin{cases} 1, & i = -1 \\ 0, & i < -1 \end{cases} \quad (1)$$

where  $\beta_0$  and  $\gamma_0$  are the on-momentum Lorentz factors. The transition gamma is defined as  $\gamma_t = \sqrt{1/\alpha_0}$ . To lowest order, all off-momentum particles have the same transition gamma when  $\alpha_1/\alpha_0 \approx -\frac{1}{2}$ , and cross transition at the same time when  $\alpha_1/\alpha_0 \approx -\frac{3}{2}$ .

For a FODO lattice with *thin* quadrupoles of strength  $B'\ell/(B\rho) = \pm S/L$ , where  $L$  is the half cell length with dipole bending angle  $\theta$ , we have [2, 3] (see also Sec.2.2.3)

$$\begin{aligned}\alpha_0 &\approx 1 - \frac{S(\hat{D}_0 - \check{D}_0)}{L\theta}, \quad \alpha_1 \approx -\frac{S(\hat{D}_1 - \check{D}_1)}{L\theta}, \\ \alpha_2 &\approx -\frac{S(\hat{D}_2 - \check{D}_2)}{L\theta} - \frac{S^3(\hat{D}_0^3 - \check{D}_0^3)}{6L^3\theta}\end{aligned}\quad (2)$$

where the dispersions at the F- and D-quadrupoles have been power expanded, respectively, as  $\hat{D} = \sum_{i=0}^{\infty} \hat{D}_i \delta^i$  and  $\check{D} = \sum_{i=0}^{\infty} \check{D}_i \delta^i$ . When  $S \ll 12$ , which is usually true because  $S = 2 \sin \frac{\mu}{2}$  and  $\mu$  is the phase advance per cell,  $\alpha_1/\alpha_0 \rightarrow +\frac{3}{2}$  and reduces to  $+\frac{1}{2}$  after chromaticities are corrected by sextupoles.

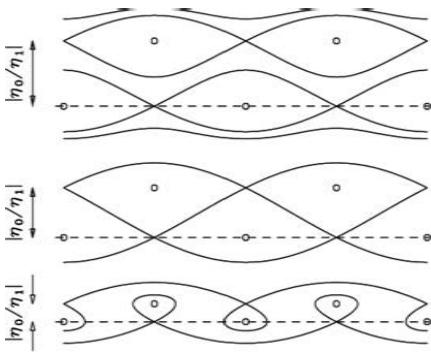
For an isochronous or quasi-isochronous ring, we must require the spread in  $\eta$  for off-momentum particles to be small also. Therefore,  $\alpha_1$  and  $\alpha_2$  need to be controlled in addition to  $\alpha_0$ . In fact, first-order effect of sextupoles alters  $\alpha_1$ , that of octupoles alters  $\alpha_2$ , etc. For example, let  $S_n = B^{(n)}\ell/(B\rho)$  be the strength of a *thin* quadrupole ( $n=1$ ), *thin* sextupole ( $n=2$ ), or *thin* octupole ( $n=3$ ) of length  $\ell$  at a location where the horizontal and vertical dispersions are, respectively  $D_x$  and  $D_y$ . Their first-order effects are [4]  $\Delta\alpha_0 = -S_1(D_x^2 - D_y^2)/C_0$ ,  $\Delta\alpha_1 = -S_2(D_x^3 - 3D_x D_y^2)/C_0$ ,  $\Delta\alpha_2 = -S_3(D_x^4 - 6D_x^2 D_y^2 + D_y^4)/C_0$ .

The Hamiltonian describing the longitudinal rf phase difference  $\Delta\phi_{\text{rf}}$  is (Sec.2.3.1) [5]

$$H = h \left( \frac{1}{2} \eta_0 \delta^2 + \frac{1}{3} \eta_1 \delta^3 + \frac{1}{4} \eta_2 \delta^4 + \dots \right) + \frac{eV_{\text{rf}}}{2\pi\beta_0^2 E_0} [\cos(\phi_s + \Delta\phi_{\text{rf}}) + \Delta\phi_{\text{rf}} \sin \phi_s] \quad (3)$$

where  $V_{\text{rf}}$  is the rf voltage with synchronous phase  $\phi_s$  and harmonic  $h$ , while  $E_0$  is the on-momentum energy. If only the  $\eta_0$  and  $\eta_1$  terms are considered, the two series of distorted pendulum-like buckets in the top figure ( $\Delta\phi_{\text{rf}}$  vs  $\delta$  with  $\phi_s = 0$  or  $\pi$ , see [2] for nonzero or non- $\pi$   $\phi_s$ ) begin to merge to the middle figure when  $|\eta_0/\eta_1|$  is lowered to

$$\left| \frac{\eta_0}{\eta_1} \right| = \sqrt{\left| \frac{6eV_{\text{rf}}}{\pi\beta_0^2 h\eta_0 E_0} \left[ \left( \frac{\pi}{2} - \phi_s \right) \sin \phi_s - \cos \phi_s \right] \right|} \quad (4)$$



With further reduction of  $|\eta_0/\eta_1|$ , the buckets become  $\alpha$ -like (lower figure), which shrink to zero when  $|\eta_0/\eta_1| = 0$ . The total bucket height  $|3\eta_0/(2\eta_1)|$  is small. It is asymmetric in momentum spread and the beam is susceptible to longitudinal head-tail instability. If the  $\eta_1$  term is eliminated, the Hamiltonian will be dominated by  $\eta_0$  and  $\eta_2$  and the bucket becomes pendulum-like again [3]. If the Hamiltonian is dominated by the  $\eta_2$  term alone, the kinetic term is similar to a quartic potential providing maximal amount of synchrotron-frequency spread and therefore Landau damping.

The first three orders of the slip factor can be extracted by measuring the synchrotron tune  $\nu_s$  while changing the rf frequency  $f_{\text{rf}}$  [2]:

$$\nu_s^2 \approx \frac{h e V_{\text{rf}} |\eta_0 \cos \phi_s|}{2\pi\beta_0^2 E_0} \left( 1 + \frac{s_1}{\eta_0} \left[ \frac{\Delta f_{\text{rf}}}{f_{\text{rf}}} \right] + \frac{s_2}{\eta_0^2} \left[ \frac{\Delta f_{\text{rf}}}{f_{\text{rf}}} \right]^2 \right)$$

$$\begin{aligned}s_1 &= -\frac{2\eta_1 - \eta_0^2}{\eta_0} + \frac{1}{\gamma_0^2} \\ s_2 &= \frac{3\eta_2\eta_0 - 2\eta_1^2}{\eta_0^2} - \frac{\eta_1}{\eta_0\gamma_0^2} + \frac{3\gamma_0^2\beta_0^2 + 2}{2\gamma_0^4}\end{aligned}\quad (5)$$

Notice that  $\Delta f_{\text{rf}}/f_{\text{rf}}$  is typically  $\mathcal{O}(\eta_0)$ .

For the application of THz near-field imaging, THz spectroscopy, and others, bunch length compressed to the order of  $\sigma_\tau \sim 1$  ps is desired. An obvious advantage is to store the bunch in the  $\alpha$ -like buckets, where the bucket half width

$$\Delta\phi_{\text{rf}} \approx |\eta_0/\eta_1| \sqrt{2\pi\beta_0^2 E_0 h |\eta_0| / (3eV_{\text{rf}} |\cos \phi_s|)}$$

is intrinsically narrow. Low- $\alpha$  operation modes have been implemented in many light sources. Essentially,  $\alpha_0$  is reduced by making the dispersion outside the achromats of the Chasman-Green lattice negative by scaling the quadrupole strengths. At BESSY II, a reduction from  $\alpha_0 = 7.3 \times 10^{-4}$

100-fold or even more is possible [6]. At SPEAR III,  $\alpha_0$  has been reduced from  $1.18 \times 10^{-3}$  240-fold. However, smaller  $\alpha_0$  implies shorter bucket height and therefore shorter beam lifetime [7]. To increase bucket height, sextupoles are used to minimize  $|\alpha_1|$ . For a more reliable operation of the machine, the low- $\alpha$  mode of BESSY II is compromised to  $\alpha_0 = 3.5 \times 10^{-5}$  with zero-current rms bunch length reduced 5-fold to  $\sigma_\tau = 3.5$  ps. For such an operation, no injection tuning of the optics is required and beam accumulates at a good rate up to a 200-bunch current of 5 mA with a 40-h lifetime. At SPEAR III, the  $\alpha/21$ -operation mode incorporates a 21-fold  $\alpha_0$ -reduction at 100 mA in 280 bunches with a 30-h lifetime, and a measured bunch length  $\sigma_\tau = 6.9$  ps. The shortest bunch length achieved has been 2.5 ps at the single bunch current 3.5  $\mu$ A, when  $\alpha_0$  is reduced 240-fold. The beam lifetime is mostly limited by Touschek effect because of the short bucket height. When the bunch length is narrow enough, beam instability often occurs due to coherent synchrotron radiation.

## References

- [1] J. Wei, PhD thesis, Stony Brook (1990, rev. 1994)
- [2] K.Y. Ng, Fermilab Reports FN-0578, FN-0852
- [3] K.Y. Ng, NIM A404 (1998) 199
- [4] D. Robin et al, PR E48, 2149 (1993)
- [5] K.R. Symon, A.M. Sessler, Proc. Int. Conf. High Energy Acc. (1956) p.44
- [6] J. Feikes et al, EPAC 04, p.2291
- [7] X. Huang et al, PAC 07, p.1308

### 2.3.12 Nonlinear Dynamics Experiments

*W. Fischer, BNL*

The goal of nonlinear dynamics experiments is to improve the understanding of single particle effects that increase the particle amplitude and lead to loss. Particle motion in storage rings is nearly conservative and for transverse dynamics the Hamiltonian in action angle variables  $(I_x, I_y, \phi_x, \phi_y)$  near an isolated resonance  $k\nu_x + l\nu_y \approx p$  is

$$H = I_x \nu_{x0} + I_y \nu_{y0} + g(I_x, I_y) + h(I_x, I_y) \cos(k\phi_x + l\phi_y - p\theta) \quad (1)$$

where  $k, l, p$  are integers,  $\theta = 2\pi s/L$  is the azimuth, and  $s$  and  $L$  are the path length and circumference respectively. The amplitude dependent tunes are given by

$$\nu_{x,y}(I_x, I_y) = \nu_{x0,y0} + \partial g(I_x, I_y)/\partial I_{x,y} \quad (2)$$

and  $h(I_x, I_y)$  is the resonance driving term (RDT). If the motion is governed by multiple resonances,  $h(I_x, I_y)$  has to be replaced by a series of terms. The particle motion is completely determined by  $g$  and  $h$ , which can be calculated from higher order multipoles (Sec.2.3.3), or obtained from simulations. Deviations from pure Hamiltonian motion occur due to synchrotron radiation damping (Sec.3.1) in lepton or very high energy hadron rings, parameter variations, and diffusion processes such as residual gas and intrabeam scattering (Sec.2.4.12). The time scale of the non-Hamiltonian process determines the applicability of the Hamiltonian analysis.

Transverse nonlinearities are introduced through sextupoles or higher order multipoles and magnetic field errors in dipoles and quadrupoles. Sextupoles can already drive all resonances. The beam-beam interaction and space charge also introduce nonlinear fields.

Intentionally introduced nonlinearities are used to extract beam on a resonance or through capture in stable islands [1] (Sec.4.12.1). Localization and minimization of nonlinearities in a ring is a general strategy to decrease emittance growth and increase the beam lifetime. The minimization of nonlinear effects can be done locally or globally. Except for resonant extraction, amplitude increase and particle loss is often the result of chaotic particle motion. Large chaotic regions allow particles to increase their amplitudes, leading to their ultimate loss. However, chaotic particles can, on average, still survive the time period of interest, i.e. the storage time.

Nonlinear dynamics experiments aim to determine either the detuning and driving terms  $g$  and  $h$  directly, or their effect on other quantities. Nonlinear phenomena observed in experiments include phase space deformations and resonant islands in Poincaré surfaces of section, nonlinear phase advances, amplitude detuning  $g$ , decoherence (Sec.2.3.10), resonance driving terms  $h$ , smear, halo formation, echoes (Sec.2.3.13), the tune response matrix [2], dynamic aperture (Sec.2.3.9), emittance growth, and particle loss. Nonlinear experiments can also be done in the longitudinal plane [3].

**Surface of section** The properties of a nonlinear Hamiltonian system can be visualized by a Poincaré surface of section, where the phase space variables of the particle trajectory are plotted turn-by-turn (TBT). This is the experimental

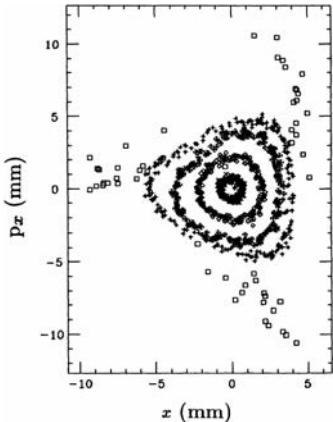


Figure 1: Experimental Poincaré surface of section near a third integer resonance obtained from BPM data after kicking the beam to different betatron amplitudes in the IUCF (reprinted with permission from [5], copyright by APS).

determination of the 1-turn map. It reveals distortions of trajectories and resonant islands. Photographs of synchrotron light from as early as 1968 show beam trapped in transverse resonance islands [4]. The surface of section can be reconstructed from two TBT readings of a pair of BPMs, where the ideal phase advance between the 2 BPMs is an odd multiple of  $\pi/2$ . For Fig.1 a 45 MeV proton beam was cooled in the IUCF and kicked to different horizontal amplitudes [5].

**Detuning and RDTs** A Hamiltonian system is completely characterized by the amplitude dependent tune shift, and the resonance driving terms. To measure the amplitude detuning, the beam is kicked to different amplitudes and the tune obtained from a spectral analysis of the TBT data [5, 6, 7, 8]. The tune error  $\Delta\nu$  in a Fast Fourier Transformation (FFT) is proportional to  $1/N$ , where  $N$  is the number of turns used, and can be improved to be proportional to  $1/N^4$  [9, 10]. Figure 2 shows a detuning measurement in VEPP-4M, where a 1.8 GeV electron beam was kicked and observed in BPMs for up to 4096 turns [6].

Resonance driving terms can also be derived from TBT BPM data (Fig.3). With an ac dipole a coherent dipole motion of indefinite length can be induced [11] and thereby the signal-to-noise ratio increased compared to a kick-based measurement [7]. The length of the TBT measurement is then limited by the BPM system.

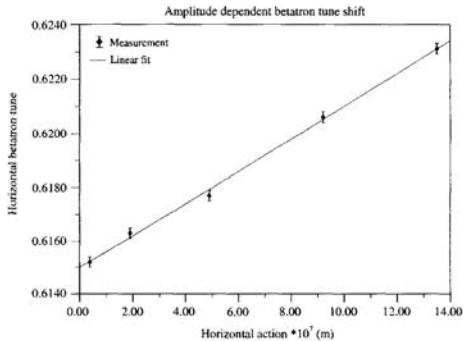


Figure 2: Typical amplitude dependence of betatron tune measured in VEPP-4M (courtesy V. Sajaev).

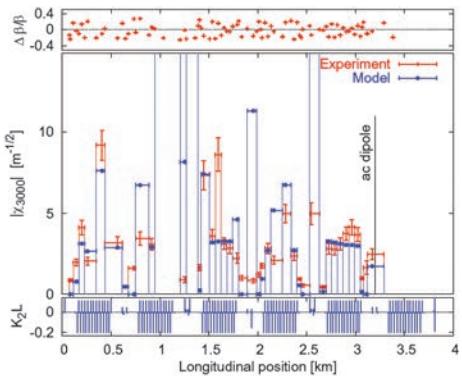


Figure 3: Measurement of sextupolar local term  $|\chi_{3000}|$  in RHIC with an ac dipole. This term is proportional to a driving term  $h$  for the resonance  $3\nu_x = p$ . The bottom plot shows the sextupolar components in the ring (reprinted with permission from [7], copyright by APS).

With the reliable measurement of  $s$ -dependent changes in resonance driving terms, multipole fields can be inferred and the correction of single or multiple resonances becomes possible [12].

**Tune and amplitude diffusion** Mapping the frequencies (tunes) as a function of initial conditions ( $I_x, I_y$ ) is often referred to as Frequency Map Analysis (FMA) [9]. Precise tune measurements [9, 10] allow the experimental determination of frequency maps that reveal potentially harmful resonances (Fig.4). Also accessible is the tune change over time  $\Delta\nu = \nu(T_1) - \nu(T_2)$ , where  $T_1$  and  $T_2$  are consecutive intervals, which is a measure of the tune diffusion.

### Sec.2.3: NONLINEAR DYNAMICS

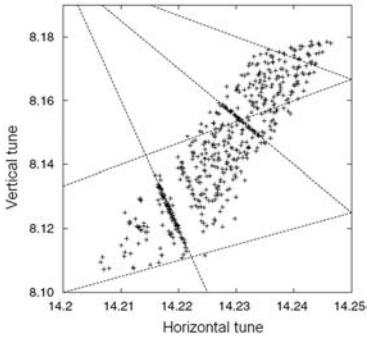


Figure 4: Experimental frequency map for the ALS. At a fixed set tune the beam is kicked to different amplitudes and the beam oscillation frequency is obtained from turn-by-turn BPM data. The dotted lines are resonances of order  $\leq 5$  (reprinted with permission from [14], copyright by APS).

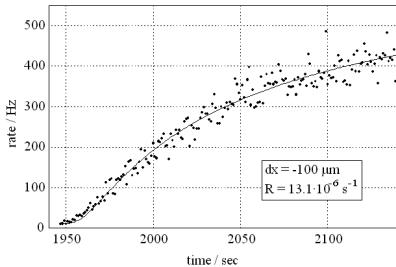


Figure 5: Particle loss rate at a HERAp collimator after retraction by  $100 \mu\text{m}$ , fitted time-dependent loss curve and diffusion constant  $R$  (courtesy M. Seidel [17]).

The time evolution of a particle distribution  $f(I, t)$  with amplitude diffusion is given by [13]

$$\frac{\partial}{\partial t} f(I, t) = \frac{\partial}{\partial I} D(I) \frac{\partial}{\partial I} f(I, t) \quad (3)$$

where  $D(I) = \langle \Delta I^2 \rangle / (2\Delta t)$  is the amplitude dependent diffusion coefficient. Over small amplitude ranges, such as those created when a scraper is moved from position  $I_c$ ,  $D(I)$  can be assumed constant and the change in the loss rate at the scraper can be fitted to obtain  $R = D(I_c)/I_c^2$  (Fig.5). Larger regions of  $I$  are sampled when the time evolution of the transverse profiles are recorded. To access large amplitudes the beam is kicked, often creating a hollow beam [8, 15]. Diffusion rates, caused by a number of mechanisms, can span many orders of magnitudes [16].

**Dynamic aperture and tune modulation** The dynamic aperture (DA) determined in simulations

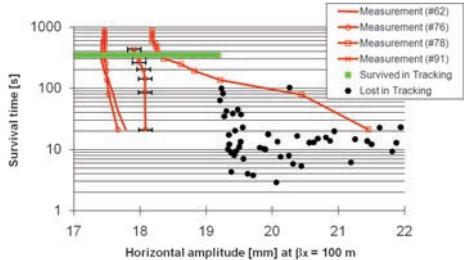


Figure 6: Measured and simulated DA in SPS with 8 strong sextupoles and tune modulation with parameters  $q = 1.87 \times 10^{-3}$  and  $\nu_M = 2.1 \times 10^{-4}$  (reprinted with permission from [8], copyright APS).

is displayed as survival plots (Sec.2.3.9). In an experiment a single large kick places a large number of particles across the DA. The DA is then determined as the maximum amplitude where particles can be observed with a transverse profile monitor, e.g. with a wire scanner (Fig.6). Increasing the emittance with many small kicks is also possible but requires a higher sensitivity in the transverse profile measurement. In lepton machines, where the survival times of particles only need to be of the same order as the synchrotron radiation damping time, the beam is usually kicked until beam loss is observed.

Tune modulation is caused by the synchrotron motion and non-zero chromaticity as well as power supply ripples in the quadrupoles. Tune modulation affects the long-term stability of particle motion. In the presence of an isolated resonance with island tune  $\nu_I$ , the modulated tune

$$\nu(N) = \nu_0 + q \sin(2\pi\nu_M N), \quad (4)$$

where  $N$  is the turn number, and  $q$  and  $\nu_M$  are the modulation depth and tune, leads to four different phases in the  $(\nu_M/\nu_I, q/\nu_I)$  diagram (Fig.7).

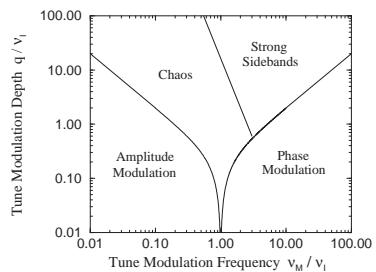


Figure 7: Approximate boundaries between dynamical phases in tune modulation space (reprinted with permission from [19], copyright by APS).

Massive chaos occurs when sidebands created by the modulation overlap (Chirikov criterion [18]). Tune modulation effects have been studied extensively [8, 19, 20].

## References

- [1] R. Cappi, M. Giovannozzi, PRST-AB 7, 024001 (2004); M. Giovannozzi et al, PRST-AB 12, 024003 (2009)
- [2] G. Franchetti, A. Parfenova, I. Hofmann, PRST-AB 11, 094001 (2008)
- [3] M. Ellison et al, PRL 70, 591 (1993)
- [4] G.N. Kulipanov et al, Novosibirsk Preprint INP 68-251 (1968)
- [5] D.D. Caussyn et al, PRA 46, 7942 (1992)
- [6] V. Kiselev et al, PA 57, 65 (1997)
- [7] R. Tomás et al, PRST-AB 8, 024001 (2005)
- [8] W. Fischer, M. Giovannozzi, F. Schmidt, PRE 55, 3507 (1997)
- [9] H.S. Dumas, J. Laskar, PRL 70, 2975 (1993)
- [10] R. Bartolini et al, PA 52, 147 (1996)
- [11] M. Bai et al, PR E56, 6002 (1997)
- [12] R. Bartolini et al, PRST-AB 11, 104002 (2008)
- [13] K.-H. Mess, M. Seidel, NIM A 351, 279 (1994)
- [14] D. Robin et al, PRL 85, 558 (2000)
- [15] T. Chen et al, PRL 68, 33 (1992)
- [16] F. Zimmermann, PA 49, 67 (1995)
- [17] M. Seidel, PhD thesis, Hamburg U., DESY 94-103 (1994)
- [18] B.V. Chirikov, Physics Report 52, 263 (1979)
- [19] T. Satogata et al, PRL 68, 1838 (1992)
- [20] O.S. Brüning, F. Willeke, PA 55, 237 (1996)

### 2.3.13 Echo Effects

*G.V. Stupakov, SLAC*

**Transverse echo** An echo effect in accelerators was first proposed in [1, 2] for the betatron motion of a beam. It could be observed in a situation where the beam in the ring is deflected off-orbit at time  $t = 0$ , causing its centroid to undergo betatron oscillations. After these oscillations have completely damped out due to beam decoherence (Sec.2.3.10), the beam is excited again by a quadrupole kick at  $t = \tau$ . This kick does not produce any visible displacement of the beam at that time, but it turns out that close to time  $t = \tau_{\text{echo}} = 2\tau$  the beam centroid undergoes transient betatron oscillations with an amplitude that is a fraction of the initial beam offset. The echo arises as a result of nonlinear beam dynamics in the phase space due to the tune dependence on the amplitude of the betatron oscillations.

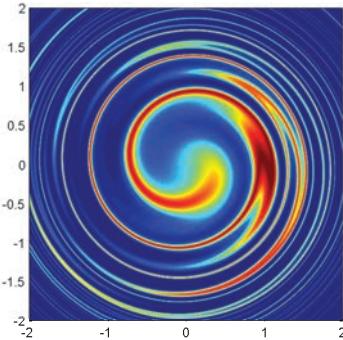


Figure 1: Transverse echo phase space ( $q, p$ ) of the beam. Note the intricate folding pattern which develops due to dependence of the rotational speed on the betatron amplitude. At a critical moment the structures in the phase line up to give a significant echo signal.

Assume that at  $t = 0$  the beam is deflected in the transverse direction by  $a$ . Define  $q \equiv x/\sqrt{\beta}$  and  $p \equiv (\alpha_x x + \beta_x x')/\sqrt{\beta}$ . Assume no  $x$ - $y$  coupling and neglect synchrotron damping and quantum excitation. The initial distribution function of the beam is assumed Gaussian,  $\psi(q, p) = (2\pi\epsilon)^{-1} \exp[-(q^2 + p^2)/2\epsilon]$ , with  $\epsilon$  the beam emittance. At time  $t = \tau$  a quadrupole kick of strength  $Q$  ( $\beta$ -function at the quad)/(focal length of the quad) is applied to the beam. We assume  $Q \ll 1$ .

Let the betatron tune be

$$\nu = \nu_0 - \mu \frac{q^2 + p^2}{\epsilon} \quad (1)$$

where  $\mu$  is the dimensionless detuning parameter. The second order perturbation theory [1, 2] predicts the echo amplitude  $\eta = \sqrt{q^2 + p^2}$ ,

$$\eta = aF\left(\frac{\tau_0}{\tau_d}, \frac{t - 2\tau}{\tau_d}\right) \quad (2)$$

where  $\tau_0 = Q\tau$ ,  $\tau_d = T_0/4\pi\mu$  is of the order of the decoherence time with  $T_0$  the revolution period, and

$$F(x, y) = \frac{x}{((1+x^2-y^2)^2+4y^2)^{3/4}} \quad (3)$$

As follows from Eqs.(2) and (3), the echo occurs at about  $t \approx 2\tau$ , and the duration of the echo pulse is of the order of the decoherence time  $\tau_d$ . It decays as  $\propto |t - 2\tau|^{-3/2}$  for large  $|t - 2\tau|$ . A characteristic phase space of the beam at the moment of the maximal echo signal is shown in Fig.1.

In addition to the echo effect at  $t = 2\tau$  higher order echos (with smaller amplitudes) can be observed at even multiples of  $\tau$  [2].

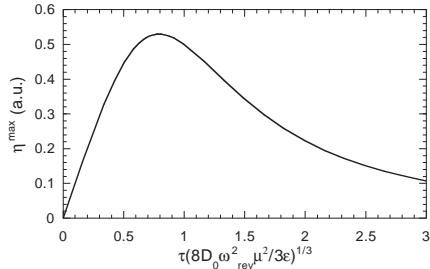


Figure 2: Echo signal (in arbitrary units) as a function of delay  $\tau$ .

Collective effects can affect the decoherence mechanism, and therefore the echo signal, significantly through an impedance in the accelerator [3].

**Effect of diffusion** The echo effect is sensitive to diffusion processes in the phase space. This is because in echo generation the phase space of the beam develops structures with characteristic size inversely proportional to the delay time  $\tau_{\text{echo}}$ , see Fig.1. These fine structures can be easily smeared out even by a weak diffusion, whose strength is amplified with diminishing dimensions of the structures. Hence, the echo can be a useful diagnostic tool for measuring weak diffusion processes in the phase space of the beam.

The effect of diffusion on echo for a bunched beam was studied in [4] when  $\tau_0 \ll \tau_d$ . In this limit the maximum value  $\eta^{\max}$  is achieved at  $t = \tau_{\text{echo}}$ ,

$$\eta^{\max} = \frac{aQ}{\tau_d} \frac{\tau}{1 + 8D_0\mu^2\omega_0^2\tau^3/3\epsilon} \quad (4)$$

where  $\omega_0 = 2\pi/T_0$  and  $D_0$  is the diffusion coefficient. See Fig.2. For small  $\tau$ , the maximum echo signal linearly increases with  $\tau$ . For large  $\tau$ , the deformation of the distribution function due to the diffusion becomes strong enough to compete with the formation of the echo, and causes  $\eta^{\max}$  to decay  $\propto \tau^{-2}$  in the limit  $\tau \rightarrow \infty$ .

**Longitudinal echo** The echo effect can also be observed in the longitudinal direction for both bunched and coasting beams. The case of a bunched beam is analogous to the transverse echo [5], with an analog of the transverse dipole kick being an instantaneous change in the rf phase, and an analog of the quadrupole kick realized as a sharp increase in the rf amplitude.

Most of the echo studies in accelerators were actually carried out with coasting beams. The theory of longitudinal echo for such beams has been

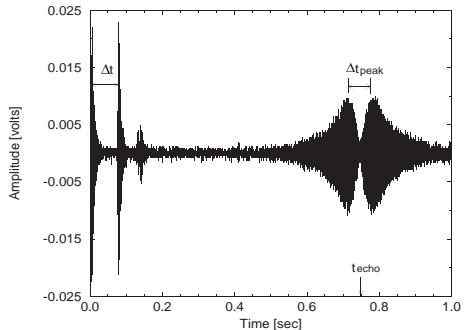


Figure 3: Beam response to impulse at  $n = 9$ , followed by  $m = 10$ .

developed in [6, 7, 8]. First, a short-duration rf pulse is applied to the beam at a revolution harmonic  $n\omega_0$ . After a delay  $\tau$ , a second pulse is applied at frequency  $m\omega_0$ . The lowest order echo can be observed at the frequency  $(n - m)\omega_0$  at a time  $\tau_{\text{echo}} = m\tau/|n - m|$ . The longitudinal echo can also be viewed as a nonlinear mixing of two waves propagating in the beam.

**Echo experiments** First echo experiments have been carried out in the Fermilab Antiproton Accumulator [6] and at CERN SPS [9]. The echo signal measured in the Fermilab experiment is shown in Fig.3. A theoretical fit to the data gave an effective diffusion rate in the beam corresponding to the diffusion time of 300 s. An extremely small diffusion rate corresponding to the doubling of the energy spread after  $10^7$  s has been measured by means of echo in the CERN experiments [10].

Transverse echo was measured in RHIC by applying a dipole kick, followed by a one-turn quadrupole kick [11].

**Using echo for seeding FEL** The echo scheme can also be used for generation of high harmonics in seeding of free electron lasers [12]. This echo-enabled harmonic generation (EEHG) mechanism offers a remarkable up-frequency conversion efficiency, with the amplitude of the generated harmonics only slowly decreasing with the harmonic number [13].

**Spin echo in accelerators** An echo occurs in the spin polarization of a beam with finite energy spread among its particles when two spin depolarization resonances are crossed (or a single resonance crossed twice). If the two crossings are spaced by a time interval  $\tau$ , the spin echo occurs at a time  $t = 2\tau$  [14].

## References

- [1] G. Stupakov, SSCL Report 579 (1992)
- [2] G. Stupakov, S. Kauffmann, SSCL Report 587 (1992)
- [3] G.V. Stupakov, A.W. Chao, PAC 1995
- [4] G. Stupakov, A.W. Chao, PAC 1997
- [5] N. Mahale et al, SSCL-N-817, 1993
- [6] L.K. Spenzouris, J.-F. Ostigy, P.L. Colestock, PRL 76, 620 (1996)
- [7] O. Brüning, CERN SL/95-83 (AP), 1995
- [8] E. Shaposhnikova, CERN SL/Note 95-125 (RF), 1995
- [9] O. Brüning et al, CERN SL-MD Note 217, 1996
- [10] O. Brüning et al, EPAC 96, 1996
- [11] W. Fischer, T. Satogata, R. Tomas, PAC 2005
- [12] G. Stupakov, PRL 102 (2009) 074801
- [13] D. Xiang, G. Stupakov, PRST-AB 12 (2009) 030702
- [14] A.W. Chao, E. Courant, PRST-AB 10 (2007) 014001

## 2.4 COLLECTIVE EFFECTS

### 2.4.1 Collective Effects in High Energy

#### Electron Linacs

*K. Kubo, K. Yokoya, KEK*

*K. Thompson, Archimedes Inc.*

**Symbols:** Average longitudinal and transverse wake functions per unit length along the beam line are  $W_{\parallel}(z)$  [unit V/C/m] and  $W_{\perp}(z)$  [unit V/C/m<sup>2</sup>] (Sec.3.2.1);  $z$  is the longitudinal separation between the exciting and the witness point charges. We only consider the monopole mode wake for longitudinal motion and the dipole mode wake for transverse motion. We ignore dependence of these functions on position  $s$  in the beam line.

#### 2.4.1.1 Single bunch longitudinal dynamics

**Energy variation and minimization**  $z$ -dependence of particle energy can be controlled by choosing the rf phase at bunch center,  $\phi_{rf}$ , along the linac. Energy spread can be also controlled by shaping the longitudinal distribution  $\rho(z)$  before injection [1].

For a short bunch, assuming sinusoidal rf voltage, off-crest phase,  $\phi_0$ , for minimum energy dependence is approximately [2]

$$\sin \phi_0 = -\frac{c}{\omega_{rf} \hat{V}} \int_{-\infty}^{\infty} \rho(z) V'(z) dz \quad (1)$$

which is proportional to the bunch charge and inverse of the bunch length. Here,  $\phi_0 > 0$  means ahead of rf crest and

$$V'(z) \equiv \frac{d}{dz} \int_z^{\infty} \rho(z') W_{\parallel}(z' - z) dz \quad (2)$$

### 2.4.1.2 Multibunch longitudinal dynamics

For multi-bunch longitudinal dynamics, only fundamental accelerating mode is important. Its effect is called beam loading.

**Symbols:**  $q$  = bunch charge,  $\omega_{rf}$  = rf frequency,  $\omega$  = resonance frequency of structure,  $\kappa_1$  = fundamental mode loss factor per length,  $T_f$  = structure filling time,  $f_v \equiv \frac{v_g}{c-v_g}$  where  $v_g$  is group velocity in the structure,  $L_s$  = structure length,  $t_b$  = bunch spacing,  $n_B$  = total number of bunches in the bunch train,  $I_b = q/t_b$  = average beam current in the bunch train. We assume  $t_b \omega_{rf}/2\pi$  = integer and  $|\omega_{rf} - \omega| \ll \omega$ .

**Standing wave structures** In a standing wave structure, beam induced voltage by a bunch train at time  $t$  after the first bunch enters the structure is [3]

$$V_b(t) = \frac{2\kappa_1 T_f I_b L_s}{1 - i \tan \psi} \left( 1 - e^{-(1-i \tan \psi) \frac{t}{T_f}} \right) \quad (3)$$

where  $\tan \psi = (\omega - \omega_{rf}) T_f$  ( $\psi$  is the detuning angle).

**Traveling wave structures** In a constant gradient structure, induced gradient at time  $t$  after the first bunch enters the structure is [3]

$$V_b(t) = \frac{r_s I_b \tau L_s}{1 - e^{-2\tau}} \left( \frac{1 - e^{-2\tau x}}{2\tau} - xe^{-2\tau} \right) \quad (4)$$

where  $r_s$  is shunt impedance per unit length ( $\Omega/m$ ),  $\tau$  the attenuation parameter,  $I = eN/t_b$  the average current,  $x = \min(t/T_f, 1)$ .  $V_b$  is constant for  $t > T_f$ .

#### Multi-bunch energy compensation methods

If the bunch train length  $\lesssim T_f$ , transient beam loading must be compensated to obtain uniform acceleration.

For standing wave structures, energy gain can be made uniform by adjusting time of rf pulse, starting rf before bunch train enters the structure by time  $t_{rf}$ ,

$$t_{rf} = T_f \ln \left( 1 + \frac{V_c}{2\kappa_1 L_s T_f I_b} \right) \quad (5)$$

where  $V_c$  is the uniform accelerating voltage for all bunches.

For SC cavities ( $Q_0 \gg Q_L$ ), almost all rf power can be given to beam (no reflection power) keeping the uniform acceleration condition, if  $t_{rf} = T_f \ln 2$ . In such case, assuming  $\psi = 0$ ,

$$V_c = -V_b(t \rightarrow \infty) = \frac{1}{2} V_g(t \rightarrow \infty) \quad (6)$$

where  $V_b(t \rightarrow \infty)$  is voltage induced by continuous beam with the same current, and  $V_g(t \rightarrow \infty)$  is voltage generated by continuous rf with the same power.

For traveling wave structures, several methods are possible.

1. Matched filling: By adjusting timing and bunch spacing of train, one may cancel most bunch-to-bunch energy variation in a short train [10]. For constant impedance structure, assuming  $n_B t_b \ll T_f$ , the first and last bunch have equal energy gains if

$$\frac{t_b}{T_f} = \frac{2\kappa_1 q / (1 + f_v)}{\sqrt{2r_s P_{rf} l_A} e^{-(1+f_v)s_0/l_A} + n_B \kappa_1 q} \quad (7)$$

Here  $s_0$  is the distance rf pulse has propagated in the structure when the first bunch is injected, and  $l_A = L/\tau$  is the attenuation length of rf. This method works well only for short bunch train.

2. Pre-filling (Sometimes called  $\Delta T$  method.): Supplying shaped rf power before beam injection can be done in such a way that the energy gain of each bunch during the transient period approximates the energy gain in steady state [11]. In the simplest form, the amplitude of the input rf is linearly ramped during the first filling time, then bunch train is injected during second filling time with constant rf amplitude. Shaping of rf wave form can be done, e.g., by controlling phases at which two or more klystron outputs are combined.

3. Staggering timing: Timing of rf pulses into a certain number of accelerating sections are distributed over filling time (or bunch train length, whichever shorter) [12, 13, 14].

4.  $\Delta f$  method: This is achieved by mixing structures of slightly different frequencies. Required total amplitude of such compensation rf field is proportional to inverse of frequency difference, while the frequency difference,  $\Delta\omega \lesssim 1/n_B t_b$ .

The last two methods have problem that compensation is nonlocal and energy spread may be too large in the low energy part of linac.

### 2.4.1.3 Single bunch transverse dynamics

The equation of motion is [6]

$$\begin{aligned} & \frac{d}{ds} \left( E(s) \frac{d}{ds} x(z, s) \right) + E(s) k^2(s) x(z, s) \\ &= e^2 \int_z^\infty W_\perp(z' - z) x(z', s) dz' \end{aligned} \quad (8)$$

The energy spread and discreteness of focusing elements are omitted.

For uniform  $\rho = N/l$  ( $l$  is length of bunch),  $W_\perp(z) = W'_\perp z$ , uniform acceleration  $E(s) = E_0 + Gs$ ,  $k(s) = k_0$ , initial condition  $x(z, 0) = x_0^0$ ,  $k_0 s \gg 1$ ,

$$x(z, s) \approx \text{Re} \left[ x_0 \sqrt{\frac{E_0}{E(s)}} e^{ik_0 s} \sum_{n=0}^{\infty} \frac{1}{n!(2n)!} \left( \frac{\eta}{2i} \right)^n \right] \quad (9)$$

where

$$\eta = \frac{e N W'_{perp} z^2 s}{k_0 (E(s) - E_0) l} \ln \left( \frac{E(s)}{E_0} \right) \quad (10)$$

For large  $\eta \gg 1$ , the sum has a asymptotic form,

$$x(z, s) \approx x_0 \sqrt{\frac{E_0}{E(s)}} \frac{\eta^{-1/6}}{6\pi} e^{\frac{3\sqrt{3}}{4}\eta^{1/3}} \quad (11)$$

This beam breakup instability is strong for linacs with large single-bunch charge and small aperture accelerating structures, such as normal conducting linear colliders. This problem is usually cured by BNS damping.

**Two-macroparticle model** With two macroparticles (spacing  $l$ , charge  $N e/2$ ) injected with initial displacement  $\hat{x}$ , solutions are given in Tab.1 [3, 4]. For the case with no acceleration ( $E = \text{constant}$ ) and constant focusing ( $k_1, k_2 = \text{constants}$ ), if  $k_1 = k_2$ ,

$$\frac{x_2^{n.a.} - x_1^{n.a.}}{\hat{x}} = \frac{N e^2 W_\perp(l)}{4kE} s \sin ks \quad (12)$$

If  $k_1 = k$  and  $k_2 = k + \Delta k$  with  $\Delta k \ll k$ ,

$$\begin{aligned} \frac{x_2^{n.a.} - x_1^{n.a.}}{\hat{x}} &\approx - \left( 1 - \frac{N e^2 W_\perp(l)}{4kE \Delta k} \right) \\ &\times 2 \sin(\Delta ks/2) \sin[(k + \Delta k/2)s] \end{aligned} \quad (13)$$

**BNS damping and autophasing** [7] In two-macroparticle model, to make head and tail macroparticles follow the same orbit, Eq.(13) requires

$$\Delta k = \frac{N e^2 W_\perp(l)}{4kE} \quad (14)$$

Table 1: Two-macroparticle model solutions.

|  |   |
|--|---|
| No acceleration  | $x_1^{n.a.} = \hat{x} \cos ks$<br>$x_2^{n.a.} = \hat{x} \left[ \cos ks + \frac{Ne^2 W_{\perp}(\ell)}{4kE} s \sin ks \right]$  |
| Uniform acceleration and smooth focusing:<br><br>$E(s) = E_0(1 + \alpha s)$<br>$k(s) = k_0 \left( \frac{E_0}{E(s)} \right)^p$<br>(adiabatic, i.e. $\alpha \ll k_0$ ) | <p>Case <math>p = 0</math>:</p> $x_1^{p=0}(s) = \hat{x} \left( \frac{E_0}{E(s)} \right)^{1/2} \cos k_0 s$ $x_2^{p=0}(s) = \hat{x} \left( \frac{E_0}{E(s)} \right)^{1/2} \left[ \cos k_0 s + \frac{Ne^2 W_{\perp}(\ell)}{4k_0 E_0 \alpha} \ln(1 + \alpha s) \sin k_0 s \right]$ <p>Case <math>p = 1/2</math>:</p> $x_1^{p=1/2}(s) = \left( \frac{E_0}{E} \right)^{1/4} x_1^{n.a.}(s_{\text{eff}}) = \left( \frac{E_0}{E} \right)^{1/4} \hat{x} \cos k_0 s_{\text{eff}}$ $x_2^{p=1/2}(s) = \left( \frac{E_0}{E} \right)^{1/4} x_2^{n.a.}(s_{\text{eff}})$ $= \left( \frac{E_0}{E} \right)^{1/4} \hat{x} \left[ \cos k_0 s_{\text{eff}} + \frac{Ne^2 W_{\perp}(\ell)}{4k_0 E} s_{\text{eff}} \sin k_0 s_{\text{eff}} \right]$ <p>where effective length: <math>s_{\text{eff}} \equiv \frac{1}{k_0} \int_0^s k(s') ds'</math></p> |

or, with  $E_1 = E$ ,  $E_2 = E + \Delta E$ ,  $\Delta E \ll E$ ,  $\frac{\Delta k}{k} = \xi \frac{\Delta E}{E}$  ( $\xi$  is related to chromaticity and usually  $\sim -1$ ),

$$\Delta E = \frac{Ne^2 W_{\perp}(l)}{4\xi k^2} \quad (15)$$

These are autophasing conditions; also sometimes called BNS damping conditions, but BNS damping may also refer to cases where betatron oscillation is actually damped in parts of the bunch.

To achieve autophasing for an arbitrary bunch distribution  $\rho(z)$  [8] (no acceleration),

$$\frac{\Delta k(z)}{k} = \frac{e^2}{2k^2 E} \int_z^{\infty} \rho(z') W_{\perp}(z - z') dz' \quad (16)$$

These autophasing conditions also apply to uniformly accelerated case with  $p = 1/2$ .

Choice of rf phase profile involves consideration of minimizing energy spread, optimization of BNS damping, and amount of sacrificed acceleration by putting bunches off rf crests. One solution, as considered in normal conducting linear colliders, is to run behind rf crests early in linac, introducing BNS energy spread, then run ahead of crests later in linac so as to reduce energy spread at the end.

**Quadrupole beam break-up** can occur if beam is sufficiently intense and transverse beam size is comparable to beam pipe radius [9]. In contrast to dipole BBU, this can occur even if beam is well steered to axis of accelerator.

#### 2.4.1.4 Multibunch transverse dynamics

The cumulative BBU, first studied at SLAC [15, 16, 17, 18], has been treated for various circumstances; long/short beam pulse, strong/weak focusing system, rf/induction/superconducting linacs, etc. [19, 20] See Sec.4.4 for the case of energy recovery linacs (ERLs).

**Equation of motion** One can treat each bunch as a point charge (since electron bunches are short). (Finite length bunch is discussed in [21].) For  $n_B$  equally-charged (bunch charge  $Ne$ ), equally-spaced (spacing  $l$ ) bunches, each bunch represented as a single macroparticle,

$$\begin{aligned} & \frac{d}{ds} \left( E(s) \frac{d}{ds} x_n(s) \right) + E(s) k_n^2(s) x_n(s) \\ &= Ne^2 \sum_{j=1}^{n-1} W_{\perp}((n-j)l) x_j(s) \end{aligned} \quad (17)$$

A difference from single-bunch BBU is that  $W_{\perp}$  is dominated by one or a few resonator modes having large shunt impedance,

$$W_{\perp}(z) = \sum_k \left( \frac{R}{Q} \right)_{\perp,k} \omega_k e^{-\frac{\omega_k z}{2cQ_k}} \sin \frac{\omega_k z}{c} \quad (18)$$

In case of a single isolated resonator with infinite  $Q$ , the amplitude blowup factor of the last bunch takes a form similar to Eq.(10) with dominant exponential factor  $e^{\sqrt{\eta}} (\eta \gg 1)$ ,

$$\eta = e^2 n_B N W_{\perp 0} \int_0^L \frac{ds}{E(s) k(s)} \quad (19)$$

## Sec.2.4: COLLECTIVE EFFECTS

where  $W_{\perp 0}$  is the coefficient of sine function for the resonator mode in Eq.(18). The condition  $\eta \lesssim 1$  crudely determines the required  $W_{\perp 0}$ .

For constant acceleration and focusing that scales as  $E^{-p}$ , solutions in WKB approximation, for an arbitrary wake, are

$$x_n(s) = \left\{ x_n(0) + \frac{Ne^2}{2iE_0 k_n(0)} \int_0^s ds' \right. \\ \times \left( \frac{E_0}{E(s')} \right)^{\frac{1-p}{2}} \exp[-i\psi_n(s', 0)] \\ \times \sum_{j=1}^{n-1} W_{\perp}((n-j)l) x_j(s') \Bigg\} \\ \times \left( \frac{E_0}{E(s)} \right)^{\frac{1-p}{2}} \exp[i\psi_n(s, 0)] \quad (20)$$

where

$$\psi_n(s, s') \equiv \int_{s'}^s k_n(s'') ds'' \quad (21)$$

These equations for  $x_n(s)$  may readily be solved numerically.

**Daisy chain model** If the wake field is negligible beyond more than one bunch spacing,

$$x_1'' + k_2 x_1 = 0 \\ x_n'' + k_2 x_n = \frac{Ne^2 W_{\perp}(l)}{E} x_{n-1} \quad (n > 1) \quad (22)$$

ignoring acceleration and variation of focusing along the linac [23]. For initial conditions  $x_n(0) = 1, x_n'(0) = 0$ ,

$$x_n(s) \approx \sum_{j=0}^{n-1} \frac{(-i\sigma s)^j}{j!} e^{iks} \quad (23)$$

where  $\sigma \equiv \frac{Ne^2 W_{\perp}(l)}{2kE}$ . For sufficiently large  $|\sigma s|$ , the amplitude of oscillation of bunch  $n$  grows as  $s^{n-1}$ . The criteria for little or no blow-up in linac is  $|\sigma L| < 1$  where  $L$  is length of linac.

Analytic daisy chain results given above still hold for case of constant acceleration and a focusing characterized by  $p = 1/2$ , provided that  $s$  is interpreted as the effective length (See Tab.1) and  $x_n(s)$  is interpreted as bunch offset with adiabatic damping  $(E_0/E)^{1/4}$  factored out.

**Control of multibunch BBU** For large scale linacs, several methods (and their combinations) have been proposed.

- Effects of wake are minimized if bunches are near zero crossing of wake, i.e. dominant deflecting mode frequency and bunch spacing satisfy  $\sin \omega_k t_b \approx 0$ .

- Effects to the next bunch are small if  $Q_s$  of dominant deflecting modes are small enough (Called “damped structure”). When wake can reach only the next bunch (daisy chain model), the condition that the blowup be small is

$$\frac{Ne^2 W_{\perp}(l)}{2} \int_0^L \frac{ds}{E(s) k(s)} < 1 \quad (24)$$

One of the most strongly damped structures is choke mode structure [25], in which  $Q$  can be as small as  $\lesssim 10$  for all deflecting modes.

- If deflecting modes have a frequency spread, the sum of the wake along a long structure (or more than one structures) can be reduced [26, 27]. Spread can come from random construction errors [28]. Or, structure can be designed so that cell dimensions differ slightly from cell to cell so as to give spread of mode frequency, causing decoherence of wake function (“detuned structure”). For sufficient reduction in one bunch spacing ( $t_b$ ), required frequency spread is  $\sigma_f \gtrsim 1/(2t_b)$ . For avoiding re-coherence at the end part of bunch train, typical difference of two frequencies next each other is required as  $\delta f \lesssim 1/(n_B t_b)$ , which requires many cells with different frequencies for long bunch train. Combining method 2 with this method (damped-detuned structure), this re-coherence effects can be mitigated [29].

### 2.4.1.5 Effects of structure misalignment

So far we discussed BBU induced by wakefield perfectly aligned wakefield sources (accelerating structures), where injection error is the seed of the oscillation. Misalignment of structures can be treated as additional effects, assuming typical misalignment is much larger than amplitude of beam oscillation. Orbit offset of particle of longitudinal position  $s$  at the end of linac (length  $L$ ) is

$$x(z, L) = -eW_{\perp, \text{sum}}(z) \int_0^L ds a(s) \frac{R_{12}}{E(s)} \quad (25)$$

where  $a(s)$  is offset misalignment of wake source at  $s$ ,  $R_{12}$  the transfer matrix from  $s$  to the end of linac,  $E(s)$  the beam energy at  $s$ , and we define “sum wake” as

$$W_{\perp, \text{sum}}(z) \equiv \int_0^z \rho(z') W_{\perp}(z - z') dz' \quad (26)$$

For multi-bunch dynamics, it may be

$$W_{\perp,\text{sum}}(n) = \sum_{j < n} eN W_{\perp}((n-j)ct_b) \quad (27)$$

Beam size increase is expressed as

$$\Delta\sigma^2(L) \approx \frac{1}{Ne} \int_0^\infty \rho(z) x^2(z, L) dz \\ = e^2 W_{\perp,\text{div}} \left( \int_0^L dsa(s) \frac{R_{12}(s)}{E(s)} \right)^2 \quad (28)$$

$R_{12}(s)$  is 1-2 element of transfer matrix from  $s$  to the end,  $E(s)$  beam energy at  $s$  and we define “divergence of sum wake” as

$$W_{\perp,\text{div}} \equiv \int_0^\infty dz \rho(z) \frac{W_{\perp,\text{sum}}^2(z)}{(Ne)^2} \\ - \left( \int_0^\infty dz \rho(z) \frac{W_{\perp,\text{sum}}(z)}{Ne} \right)^2 \quad (29)$$

For random and independent misalignment of each structure of length  $L_s$  (this may be replaced by length of independently aligned unit),

$$\left\langle \left( \int_0^L dsa(s) \frac{R_{12}}{E(s)} \right)^2 \right\rangle = L_s^2 \sum_i \langle a_i^2 \rangle \left( \frac{R_{12,i}}{E_i} \right)^2 \quad (30)$$

where  $i$  is index of structure, and  $s$  dependence is replaced by  $i$  dependence for every term. Increase of projected, normalized emittance is evaluated from  $\Delta\sigma^2(L)$  dividing by beta-function and multiplying energy factor at the end,  $\Delta\gamma\epsilon = \gamma(L)\Delta\sigma^2(L)/\beta(L)$ . Expressions above are suitable for numerical calculation.

For a long linac, one may transform  $\langle a_i^2 \rangle \rightarrow a^2$  ( $a$  is rms misalignment),  $\sum_i \rightarrow \frac{L}{L_s}$ , and  $\left( \frac{R_{12,i}}{E_i} \right)^2 \rightarrow \beta(L)/2(\overline{\beta/E})E(L)$  ( $\overline{\cdot}$  denotes average along linac), then

$$\Delta\gamma\epsilon = \frac{e^2}{2mc^2} a^2 L_s L W_{\perp,\text{div}} \overline{\left( \frac{\beta}{E} \right)} \quad (31)$$

For multibunch beam of very long pulse, since the wake function eventually dies away as  $z \rightarrow \infty$  due to finite  $Q$ ,  $W_{\perp,\text{sum}}(n) \approx \text{constant}$  for large  $n$ , bunches in tail part of the beam have the same orbit [22].

## References

- [1] F.J. Decker, R. Holzapple, T. Raubenheimer, LINAC94 (1994) p.47
- [2] K. Bane, AIP Proc. 153(1987)
- [3] P. Wilson, AIP Proc. 87(1981)
- [4] A. Chao, Physics of Collective Instabilities in High Energy Accelerators, Wiley (1993)
- [5] A. Piwinski, DESY report HERA 92-11(1992)
- [6] A.W. Chao, B. Richter, C.Y. Yao, NIM 178 (1980)
- [7] V. Balakin, S. Novokhatsky, V. Smirnov, Proc. 12th Int. Conf. on High Energy Acc. (1983)
- [8] V. Balakin, Proc. Workshop on Linear Colliders (1988) p.55
- [9] A. Chao, R.K. Cooper, PA 13(1083)1
- [10] R.D. Ruth, Proc. ICFA/INFN Workshop on Physics of Linear Colliders (1988)
- [11] K.A. Thompson, R.D. Ruth, PAC93
- [12] R.B. Neal, ed., The Stanford Two-Mile Accelerator, Benjamin (1968) p.84
- [13] R.H. Helm et al, IEEE Trans. Nucl. Sci. NS-16, No.3(1969)311
- [14] B. Aune, J. Leroy, A. Mosnier, PAC83, p.2995
- [15] R.B. Neal, W.H. Panofsky, Science (1966)1353
- [16] R.H. Helm, G.A. Loew, Linear Accelerators, North-Holland (1970)
- [17] W.K.H. Panofsky, M. Bander, RSI 39 (1968)206
- [18] V.K. Neil, L.S. Hall, R.K. Cooper, PA 9 (1979) 213
- [19] Y.Y. Lau, PRL 63 (1989) 1141
- [20] G.H. Hoffstatter, I.V. Bazarov, PRST-AB 7 (2004) 054401
- [21] C.L. Bohn, J.R. Delayen, PR A45 (1992) 5964
- [22] R.L. Gluckstern, R.K. Cooper, P.J. Channell, PA 16 (1985) 125
- [23] K.A. Thompson, R.D. Ruth, PR D41 (1989) 964
- [24] G.F. Caparaso, Linac86, 1986, p.17; G.F. Caparaso et al, PRL 57 (1986) 1591
- [25] T. Shintake, Jap. J. Appl. Phys. 31, L1567 (1992)
- [26] K. Yokoya, DESY 86-084 (2986)
- [27] R.L. Gluckstern, F. Neri, R.K. Cooper, LINAC86, p.540
- [28] R.L. Gluckstern, F. Neri, R.K. Cooper, PA 23 (1988) 37
- [29] K.A. Thompson et al, PA 47 (1994) 65

## 2.4.2 Collective Effects in Energy Recovery Linacs

*G. Hoffstaetter, Cornell U.*

Energy Recovery Linacs (ERLs) accelerate in a recirculating linac with  $N_p$  passes to high energy where the beam is used for experiments. Subsequently the spent beam is decelerated through  $N_p$  passes to low energy where it is dumped. ERLs are therefore nonperiodic systems and all collective effects of one-pass systems (Sec.2.4.1) can be relevant, e.g. single-bunch beam breakup,

accumulative beam breakup, space charge forces, coherent synchrotron radiation, and ion trapping.

While linac beam currents are usually limited by the available power, this limitation is lifted when the energy of the spent beam is recovered, and much larger currents become available. Some collective effects common in ring accelerators therefore can become relevant, e.g. Touschek scattering, intrabeam scattering, fast ion instability and electron cloud.

However, there is a class of instabilities that is pertinent to ERLs and specific to the nature of multi-turn, nonperiodic systems. This is the transverse and longitudinal recirculatory beam-breakup instability and the recirculatory orbit distortion.

For currents higher than a threshold  $I_{th}$ , the power in HOMs can resonantly increase. The HOM indexed with subscript  $\lambda$  has angular frequency  $\omega_\lambda$ , quality factor  $Q_\lambda$ , impedance  $(R/Q)_\lambda^{[2n]} Q_\lambda$  (in circuit definition,  $\frac{1}{2}$  of linac definition, and units of  $\frac{\Omega}{m^{2n}}$ ) for a HOM of multipole order  $n$ , polarization angle  $\theta_\lambda$  and polarization vector  $\vec{e}_\lambda = \cos \theta_\lambda \vec{e}_x + \sin \theta_\lambda \vec{e}_y$ . The damping time is  $\tau_\lambda = 2 \frac{Q_\lambda}{\omega_\lambda}$ . The bunch repetition rate is  $t_b$ . The wake potential (Sec.3.2.1) is

$$W(t) = c \left( \frac{R}{Q} \right)_\lambda^{[2n]} e^{-\frac{t}{\tau_\lambda}} \sin \omega_\lambda t \quad (1)$$

A general formulation of recirculatory BBU for dipole modes can be found in [1] for uncoupled, and in [2] for coupled optics. A general theory for quadrupole and  $2n$ -pole modes is outlined in [3]. For 1 degree of freedom, [4] has a theory of BBU instability in recirculating linacs, i.e. energy is added in each pass, and where every bucket is filled; [5] adds subharmonic bunching. Tracking simulations [6] compared well with this theory. Coupling and polarized modes were considered for one HOM and one turn in first order theory [8]. Comparisons between experiments and theory have been successful, including polarized modes and coupled optics, e.g. in [9]. Longitudinal recirculatory BBU is analyzed in [10].

### One dipole HOM and one recirculation

The return time is  $t_r$  and the linear beam optics in the transverse plain is described by  $\Delta \vec{x} = \mathbf{M} \Delta \vec{x}'$ , so that the excitation of HOM  $\mu$  by a kick in HOM  $\nu$  is characterized by  $D_{\mu\nu} = \vec{e}_\mu^T \mathbf{M} \vec{e}_\nu$  and for one mode  $D = D_{\lambda\lambda}$ . For horizontal HOMs this reduces to  $D = \sqrt{\frac{E_1}{E_2}} \sqrt{\beta_1 \beta_2} \sin \Psi$ , where the index 1 refers to the particle after the HOM in the

first turn, and 2 indicates the return pass, with the energies  $E_i$ , the transverse beta functions  $\beta_i$ , and the betatron phase advance  $\Psi$  between 1 and 2.

If there is only one HOM and one recirculation turn, the threshold current can be determined by the following three cases:

$$I_{min} = \frac{2E_1}{ec} \frac{1}{\tau_\lambda \left( \frac{R}{Q} \right)_\lambda^{[2n]}} \frac{1}{|D \sin \Psi_{12}|} \quad (2)$$

$$t_r \ll \tau_\lambda, D \sin(\omega_\lambda t_r) < 0 : I_{th} = \frac{I_{min}}{|\sin(\omega_\lambda t_r)|}$$

$$t_r \ll \tau_\lambda, D \sin(\omega_\lambda t_r) > 0 :$$

$$I_{th} = I_{min} \sqrt{1 + \left[ \frac{\tau_\lambda}{t_r} \text{mod}(\omega_\lambda t_r, \pi) \right]^2}$$

$$t_r \gg \tau_\lambda :$$

$$I_{th} = I_{min} \sqrt{1 + \left[ \frac{\tau_\lambda}{t_r} \text{mod}(\omega_\lambda t_r \pm \frac{\pi}{2}, 2\pi) \right]^2}$$

Occasionally, additional factors are found in the literature, e.g. [7], notably  $e^{-\frac{t_b}{\tau_\lambda}}$ . In [1] it has been shown that such additional factors are not required.

**General transverse recirculatory BBU** A general theory for  $N_p$  passes and  $N$  HOMs can be formulated in the form

$$\lambda_k = k^{\text{th}} \text{ eigenvalue of } [\mathbf{W}(\omega) \mathbf{U}] \quad (3)$$

and the threshold current is the smallest real  $1/\lambda_k$  for  $\omega \in [0, \frac{\pi}{t_b}]$ . The matrix  $\mathbf{W}(\omega)$  has  $N_p^2 N^2$  coefficients  $W_{ik}^{LI}$  that describes how the  $i^{\text{th}}$  HOM in the  $L^{\text{th}}$  turn accumulates fields from orbit oscillations excited by the  $k^{\text{th}}$  HOM in the  $I^{\text{th}}$  turn, and  $U_{kj}^{IJ}$  describes how transverse momenta changed at the HOM for  $j$  and  $J$  produce orbit offsets at the  $k^{\text{th}}$  HOM in turn  $I$ .  $\mathbf{U}$  therefore depends on coupled motion and HOM polarization angles. The form of  $\mathbf{W}(\omega)$  and  $\mathbf{U}$  can be found in [1].

The growth rate  $\alpha$  can be obtained by

$$\alpha = -\frac{1}{I_{th}^2} \text{Im}\left\{ \left( \frac{d\lambda_k}{d\omega} \right)^{-1} \right\} \quad (4)$$

evaluated for the  $k$  and the  $\omega$  in Eq.(3) that lead to the threshold current.

### Transverse higher order recirculatory BBU

A multipole of order  $n$  produces the transverse kick  $f_x + i f_y = n(x - iy)^m e^{in\theta_\lambda}$  with  $m = n - 1$  and the HOM excitation [3]

$$D = \langle (f_x, f_y)(\vec{x}_2) \mathbf{M} \begin{pmatrix} f_x \\ f_y \end{pmatrix} (\vec{x}_1) \rangle_1 \quad (5)$$

where  $\vec{x}_2$  is expressed in terms of  $\vec{x}_1$ , and the average is performed over the phase space at location 1. For dipole HOMs this amounts to  $D$  in Eq.(2).

For round beams with decoupled and rotationally symmetric transport,

$$D = M_{12} \langle n^2 \operatorname{Re}\{[(x_2 - iy_2)(x_1 + iy_1)]^m\} \rangle_1$$

$$D = 4 \frac{E_1}{E_2} \beta_1 \beta_2 \epsilon_x \sin(2\Psi) \text{ for quad modes,}$$

$$D = 3^2 8 \left( \frac{E_1}{E_2} \beta_1 \beta_2 \right)^{\frac{3}{2}} \epsilon_x^2 \sin \Psi \cos^2 \Psi$$

for sext modes,

$$D(n\Psi) = n^2 m! \left( \frac{E_1}{E_2} \beta_1 \beta_2 \right)^{\frac{n}{2}} \epsilon_x^m \sin(n\Psi) \quad (6)$$

The last line for a general  $2n$ -pole shows only the term with leading betatron frequency  $n\Psi$  as an order of magnitude estimate. These  $D$  are inserted into Eq.(2) to determine the threshold current in the applicable regimes.

**Longitudinal recirculative BBU** In the simplest model, monopole modes drive energy oscillations above the threshold current

$$I_{th} = \frac{2E_1}{\eta t_r \omega_\lambda (\frac{R}{Q})_\lambda Q_\lambda} \quad (7)$$

with the slip factor  $\eta$  describing the relative time delay per relative energy difference for one pass.

**Recirculative orbit distortion** When the  $i^{th}$  HOM is offset by  $x_{0i}$  and a current  $I_0$  is accelerated, the orbit distortions  $x_i - x_{0i}$  in equilibrium are given by

$$\vec{x} - \vec{x}_0 = [I_0 \mathbf{W}(0) \mathbf{U} - \mathbf{1}]^{-1} \vec{x}_0 \quad (8)$$

As shown in [1], the matrix is invertible for any  $I_0$  smaller than the dipole BBU threshold current  $I_{th}$ .

## References

- [1] G.H. Hoffstaetter, I. V. Bazarov, PRST-AB 7, 054401 (2004)
- [2] G.H. Hoffstaetter, I.V. Bazarov, C. Song, PRST-AB 10, 044401 (2007)
- [3] C. Song, G.H. Hoffstaetter, Report Cornell-ERL-07-10 (2007)
- [4] J.J. Bisognano, R.L. Gluckstern, PAC87, p.1078
- [5] B.C. Yunn, PAC91, p.1785 1787 (1991)
- [6] G.A. Kraft, J.J. Bisognano, PAC87, p.1356
- [7] B.C. Yunn, PRST-AB 8, 104401 (2005)
- [8] E. Pozdnyev, PRST-AB 8, 054401 (2005)
- [9] D.R. Douglas et al, PRST-AB 9, 064403 (2006)
- [10] J.J. Bisognano, M.L. Fripp, CEBAF-PR-89-018 (1989)

## 2.4.3 Beam Loading

*D. Boussard, CERN*

### 2.4.3.1 Single-bunch passage in a cavity

The amplitude of beam-induced voltage is

$$V_{b0} = \frac{q}{C} = \frac{R}{Q} \omega_c q \quad (1)$$

where  $q$  = single bunch charge (assumed short bunch with  $\sigma_z \ll \lambda_{rf}$ ),  $R/Q$  = cavity geometric parameter,  $C = (\omega_c R/Q)^{-1}$  = equivalent shunt capacitance of cavity,  $f_c = \omega_c/2\pi$  = cavity resonant frequency.

Energy deposited by the bunch in the cavity is

$$W_b = \frac{1}{2} \frac{q^2}{C} = \frac{1}{2} \frac{R}{Q} \omega_c q^2 = \frac{1}{2} q V_{b0} \quad (2)$$

The net energy received by the bunch is

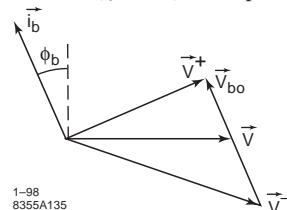
$$W = q V_g \sin \phi_b - W_b = q \left( V_g \sin \phi_b - \frac{1}{2} V_{b0} \right) \quad (3)$$

where  $V_g$  = generator voltage, and  $\phi_b$  = beam phase angle (counted from zero crossing of rf wave). The net voltage seen by the bunch is

$$V = V_g \sin \phi_b - \frac{V_{b0}}{2} \quad (4)$$

Observations: (i) The bunch sees only half of its own induced voltage ("Fundamental Theorem of Beam Loading" [1]), and (ii) Loss factor of cavity on a particular mode is  $\frac{1}{2} \frac{1}{C} = \frac{1}{2} \frac{R}{Q} \omega_c$ .

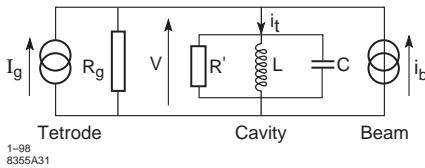
A phasor diagram at the rf frequency is shown below with  $\vec{V}^-$  and  $\vec{V}^+$  = rf voltages before and after the bunch passage,  $\vec{V}$  = net voltage seen by the bunch, and  $\vec{i}_b$  = rf component of bunch current. The vectors  $\vec{V}_{b0}$  and  $\vec{i}_b$  are in phase.



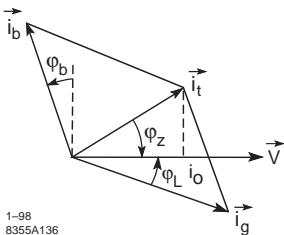
### 2.4.3.2 Cavity equivalent circuit [2]

In the figure below,  $\vec{V}$  = cavity voltage,  $R_g$  = shunt resistance of tetrode seen from cavity (transformed to the gap),  $\vec{i}_g$  = pure current source (transformed to the gap), and  $\vec{i}_b$  = rf component

of beam current. Note that maximum power and efficiency of the tetrode does not correspond to  $R_g = R'$ .



The phasor diagram (steady state, above transition case) is shown below.



Notation:  $i_0 = V/R$  = generator current needed to produce  $V$  when the cavity is at resonance and there is no beam current,  $R = R'R_g/(R' + R_g)$ ,  $Y = i_b/i_0 = V_b/V$  = beam loading parameter,  $V_b$  = beam-induced voltage at resonance,  $\phi_b$  = stable phase angle (short bunches) measured from zero crossing of rf wave,  $\phi_z$  = cavity impedance phase angle ( $\tan \phi_z = \Delta\omega/\sigma$ ,  $\Delta\omega = \omega_c - \omega_{rf}$ , and  $\sigma = \omega_c/2Q_L$ ),  $Q_L = R/(R/Q)$  = loaded quality factor of the cavity, and  $\phi_L$  = phase angle between generator current (or grid voltage) and  $\tilde{V}$ .

**Steady state equations** From the phasor diagram,

$$\tan \phi_L = \frac{\tan \phi_z - Y \cos \phi_b}{1 + Y \sin \phi_b} \quad (5)$$

$$i_g = \frac{i_0(1 + Y \sin \phi_b)}{\cos \phi_L} \quad (6)$$

Optimum detuning (real load) is obtained when  $\phi_L = 0$ , or

$$\tan \phi_z = Y \cos \phi_b, \text{ or } \Delta\omega = \omega_{DO} = Y \sigma \cos \phi_b \quad (7)$$

This is the usual operating condition of a cavity which is automatically obtained by its servo tuner.

**Case of a matched generator (e.g. klystron and circulator)** Use the same equivalent circuit, where  $R_g$  is the transmission line impedance (transformed to the gap) and replace  $\vec{i}_g$  by  $2\vec{i}_1$ , where  $\vec{i}_1$  = forward traveling wave current in line

(transformed to the gap). The cavity coupling coefficient  $\beta_c$  is given by  $\beta_c = R'/R_g$ .

The forward power  $P_f$  in the line (i.e. power flowing from klystron + circulator towards cavity) is given by

$$P_f = \frac{1}{2} R_g i_1^2 = \frac{1}{8} R_g i_g^2 \quad (8)$$

At optimum coupling,  $\beta_c = \beta_{\text{opt}}$ , defined as

$$\beta_{\text{opt}} = \frac{R'}{R_c} = 1 + \frac{R' i_b \sin \phi_b}{V} \quad (9)$$

and optimum detuning  $\Delta\omega = \omega_{DO}$ , the combination (cavity + beam) is matched to the generator and the required rf power is minimum.

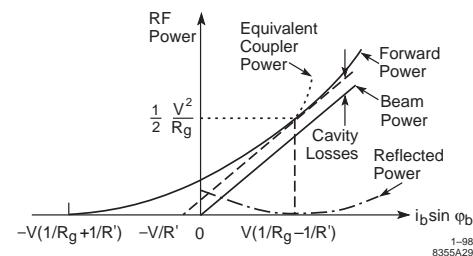
$$P_f = P_{f\min} = \frac{1}{2} \frac{V^2}{R_g} = \frac{1}{2} \frac{V^2}{R'} + \frac{1}{2} V i_b \sin \phi_b \quad (10)$$

where the first term on the r.h.s is cavity loss, while the second term is the power transferred to the beam. There is no reflected power dissipated in the circulator load (see figure below).

**Equivalent coupler power**  $P_{\text{eq}}$  is the traveling-wave power in the line which produces the same peak fields. It is an important parameter for the design of rf couplers, usually tested in traveling wave conditions. For a (cavity + beam) undercoupled i.e.  $\beta_c < \beta_{\text{opt}}$ ,  $P_{\text{eq}}$  is independent of  $i_b$ ,

$$P_{\text{eq}} = \frac{1}{2} \frac{V^2}{R_q} \quad (11)$$

The following figure shows the rf powers vs beam current for a cavity driven by a matched generator and optimum detuning  $\Delta\omega = \omega_{D0}$  or  $\phi_L = 0$ :



### **2.4.3.3 Transmission of small modulations (AM and PM) through a cavity with beam loading [2]**

To evaluate the dynamic behaviour of an rf system including feedback loops (amplitude, phase,

tuning, etc.), it is useful to know the transmission of small modulations around the static operating point. The inputs are modulations of  $i_g$  (generator current) in amplitude (AM) and phase (PM), modulations of  $i_b$  (RF component of beam current), and tune modulations. The outputs are the amplitude and phase modulations of the cavity voltage  $V$ .

#### **Generator current $i_g$ to cavity voltage $V$**

PM to PM

$$G_{pp}^G = \frac{1}{D} \left[ \sigma^2(1 + \tan^2 \phi_z) + \sigma(1 + Y \sin \phi_b)s + \sigma^2 Y (\sin \phi_b - \tan \phi_z \cos \phi_b) \right] \quad (12)$$

PM to AM

$$G_{pa}^G = \frac{1}{D} \left[ -\sigma^2 Y (\cos \phi_b + \tan \phi_z \sin \phi_b) + \sigma(\tan \phi_z - Y \cos \phi_b)s \right] \quad (13)$$

AM to AM

$$G_{aa}^G = G_{pp}^G \quad (14)$$

AM to PM

$$G_{ap}^G = -G_{pa}^G \quad (15)$$

where  $D = s^2 + 2\sigma s + \sigma^2(1 + \tan^2 \phi_z)$  with  $s$  the Laplace variable.

#### **Beam current $i_b$ to cavity voltage $V$**

PM to PM

$$G_{pp}^B = \frac{Y}{D} \left[ \sigma^2(\tan \phi_z \cos \phi_b - \sin \phi_b) - \sigma \sin \phi_b s \right] \quad (16)$$

PM to AM

$$G_{pa}^B = \frac{Y}{D} \left[ \sigma^2(\tan \phi_z \sin \phi_b + \cos \phi_b s) + \sigma \cos \phi_b s \right] \quad (17)$$

AM to AM

$$G_{aa}^B = G_{pp}^B \quad (18)$$

AM to PM

$$G_{ap}^B = -G_{pa}^B \quad (19)$$

#### **Tune modulations from $x = \Delta\omega/\sigma$ to cavity voltage $V$**

PM

$$G_{xp} = \frac{\sigma^2 + \sigma s}{D} \quad (20)$$

AM

$$G_{xa} = \frac{-\sigma^2 \tan^2 \phi_z}{D} \quad (21)$$

#### **2.4.3.4 Periodic beam loading at multiples of $f_0$ [3, 4]**

Modulated beam current (e.g. unequal filling of bunches) at multiples of the revolution frequency  $f_0$  results in AM and PM of  $V$ . Self-consistent solution, valid for only one cavity in the ring or identical, equal-voltage cavities, is given by

AM

$$\frac{a_v}{a_b} = -\omega_{DO} \times \frac{\omega_{DO}(1 + \tan^2 \phi_b) - \Delta\omega + (s + \sigma) \tan \phi_b}{(s + \sigma)^2 - \Delta\omega[\omega_{DO}(1 + \tan^2 \phi_b) - \Delta\omega]} \quad (22)$$

PM

$$\frac{P_v}{a_b} = -\omega_{DO} \times \frac{[\omega_{DO}(1 + \tan^2 \phi_b) - \Delta\omega] \tan \phi_b - (s + \sigma)}{(s + \sigma)^2 - \Delta\omega[\omega_{DO}(1 + \tan^2 \phi_b) - \Delta\omega]} \quad (23)$$

**Optimal detuning** For optimal detuning with  $\Delta\omega = \omega_{DO} = \sigma Y \cos \phi_b$ , which corresponds to a real load to the generator, one has

AM

$$\frac{a_v}{a_b} = -\frac{\omega_{DO}(s + \sigma + \omega_{DO} \tan \phi_b) \tan \phi_b}{(s + \sigma)^2 - \omega_{DO}^2 \tan^2 \phi_b} \quad (24)$$

PM

$$\frac{P_v}{a_b} = -\frac{\omega_{DO}(s + \sigma - \omega_{DO} \tan^3 \phi_b)}{(s + \sigma)^2 - \omega_{DO}^2 \tan^2 \phi_b} \quad (25)$$

**Magic detuning** For “magic” detuning with  $\Delta\omega = \omega_{DO}(1 + \tan^2 \phi_b)$ , one obtains simple, first order responses

AM

$$\frac{a_v}{a_b} = -\frac{\omega_{DO} \tan \phi_b}{s + \sigma} \quad (26)$$

PM

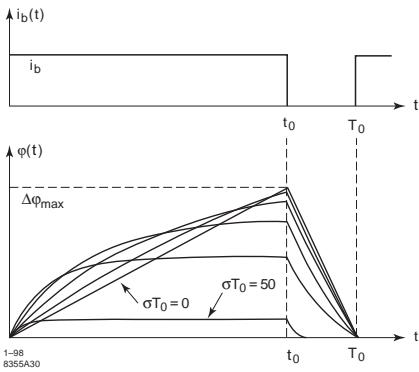
$$\frac{P_v}{a_b} = -\frac{\omega_{DO}}{s + \sigma} \quad (27)$$

With magic detuning, perfect compensation of phase modulations due to beam gaps in collider rings is possible [3].

**Uniform bunch train** For a uniform bunch train of length  $t_0$  followed by a gap and  $\Delta\omega = \omega_{DO}$ ,  $\phi_b \approx 0$ , there is no amplitude modulation, and the phase modulation corresponds to a single first-order response with maximum phase excursion (for  $\sigma \ll 1/T_0$ )

$$\Delta\phi_{\max} = \frac{1}{2} \frac{R}{Q} \frac{\omega_c}{V} \bar{i}_b (T_0 - t_0) = \omega_{DO}(T_0 - t_0) \quad (28)$$

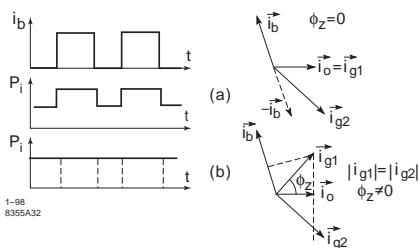
where  $\bar{i}_b$  = average rf component of beam current,  $T_0$  = revolution period. Periodic modulation of beam phase (w.r.t. an unmodulated RF) follows exactly the phase modulation of  $\vec{V}$ , as the energy gain per turn must be identical for all bunches. This is shown below for several values of  $\sigma T_0$ .



#### 2.4.3.5 Rf power needed for transient beam-loading correction [4]

The objective is to keep  $\vec{V}$  constant independent of  $\vec{i}_b$ . The solution is to provide via the rf generator an additional current  $\Delta \vec{i}_g = -\vec{i}_b$  to cancel out  $\vec{i}_b$  variations which cannot be corrected by cavity detuning (usually slow).

For a fully modulated beam (e.g. injection transient, beam gap), the required power is also generally modulated ( $\vec{i}_g = \vec{i}_{g1}$  or  $\vec{i}_{g2}$ ), e.g. if the cavity is tuned for no beam [ $\Delta\omega = 0$ ,  $\phi_z = 0$ , case (a) of figure below]. However, for a half-detuned cavity [case (b)], the power is unmodulated (the required peak power is in this case minimum).



**Installed peak power for a tetrode ( $\phi_b = 0$ )**  
Consider two cases: (a) cavity on tune for zero current, and (b) cavity half detuned.

Case a

$$P_i = \frac{1}{2} V \sqrt{i_0^2 + i_b^2} \approx \frac{V i_b}{2} \quad \text{for } i_b \gg i_0 \quad (29)$$

Case b

$$\Delta\omega = \frac{1}{4} \frac{R}{Q} \omega_c \frac{i_b}{V}$$

$$P_i = \frac{1}{2} V \sqrt{i_0^2 + \left(\frac{i_b}{2}\right)^2} \approx \frac{V i_b}{4} \quad (30)$$

for  $i_b \ll i_0$

**Installed peak power for a klystron and circulator ( $\phi_b = 0$ )**

$$P_i = \frac{R_g}{8} \left[ V^2 \left( \frac{1}{R_g} + \frac{1}{R'} \right)^2 + \left( \frac{V}{x} - i_b(t) \right)^2 \right] \quad (31)$$

where  $x = \frac{1}{2} \frac{R}{Q} \frac{\omega_c}{\Delta\omega}$  = cavity reactance.

With cavity on tune for zero beam current ( $\Delta\omega = 0$ ) and optimum matching, one has for case (a),

$$\frac{1}{R_g^2} = \left( \frac{i_b}{V} \right)^2 + \frac{1}{R'^2} \quad (32)$$

$$P_i = \frac{V^2}{4R'} + \frac{V^2}{4R_g} \approx \frac{V i_b}{4} \quad \text{for } R' i_b \gg V \quad (33)$$

$P_i$  fluctuates between  $V i_b / 8$  (when  $i_b = 0$ ) and  $V i_b / 4$  (when  $i_b = i_b \max$ ).

With cavity half detuned,  $\Delta\omega = (1/4)(R/Q)(\omega_c i_b / V)$ , and optimum matching, one has for case (b),

$$\frac{1}{R_g^2} = \left( \frac{i_b}{2V} \right)^2 + \frac{1}{R'^2} \quad (34)$$

$$P_i = \text{constant} = \frac{V^2}{4R'} + \frac{V^2}{4R_g} \approx \frac{V i_b}{8} \quad (35)$$

for  $R' i_b \gg V$

#### 2.4.3.6 Traveling-wave cavities [5, 6]

In smooth approximation the transient response of a matched traveling-wave cavity to a point charge having synchronous velocity is an rf burst starting at time  $z/v_p$  and of duration  $z/v_g$  ( $v_p, v_g$  = phase and group velocities in the structure,  $z$  = longitudinal coordinate).

In frequency domain,

$$E_b(z, \omega) = i_b \frac{R_2}{2} e^{-i\omega z/v_p} \frac{1 - e^{-i\tau z/L}}{i\tau/L} \quad (36)$$

where  $E_b$  = beam-induced rf field,  $R_2$  = effective shunt impedance of the structure ( $R_2 = E^2/2P$ ),  $L$  = total length of structure,  $\beta$  = relativistic factor, and  $\tau$  = transit time factor,

$$\tau = L \frac{\Delta\omega}{v_g} \left( 1 - \frac{v_g}{\beta c} \right) \quad (37)$$

The beam-induced voltage is

$$V_b = \frac{R_2 L^2}{4} i_b \left[ \left( \frac{\sin(\tau/2)}{\tau/2} \right)^2 - 2i \frac{\tau - \sin \tau}{\tau^2} \right] \quad (38)$$

For traveling-wave cavities the transfer impedances from generator and beam are different. This is in contrast with standing wave cavities, which for a given mode exhibit the same transfer function (circle in figure below) from generator or from beam to cavity voltage. Perfect correction of beam loading is not possible with traveling-wave structures. This is illustrated by the transient response of a traveling-wave cavity to a step in generator power (linear rise of voltage) and to a step in beam current (parabolic rise of voltage).

Figure below shows the impedance seen by the beam. Solid curve is for traveling-wave structure. Dotted curve is for standing-wave structure.

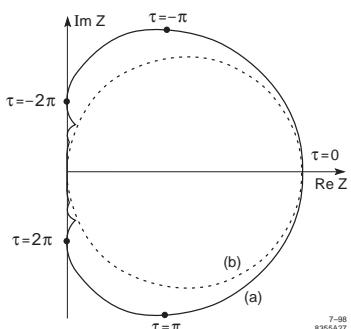
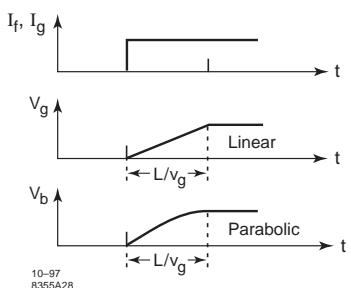


Figure below shows the transient response of a traveling-wave structure.



## References

- [1] P. Wilson, IXth Int. Conf. High Energy Acc. (1974) p.57
- [2] F. Pedersen, PAC 75, p.1906
- [3] F. Pedersen, CERN/PS 92-59 (RF)
- [4] D. Boussard, CERN SL/91-16 (RFS)
- [5] G. Dôme, Proton Linac Conf. (1976) p.138
- [6] D. Boussard, CERN Acc. School (1996)

## 2.4.4 Space-Charge Dominated Beams in Guns and Transport Lines

*M. Ferrario, INFN-LNF*

**Laminarity parameter** The dynamics of high peak current low emittance beams is strongly dominated by space charge effects [1]. Electron emission and transport is affected by self-fields produced by the electron bunch itself and by the image charge induced on metallic boundaries, which oppose the applied fields. A beam characterised by an rms envelope  $\sigma = \sqrt{\langle x^2 \rangle}$  and transverse normalised rms emittance

$$\varepsilon_n = \langle \gamma \rangle \varepsilon_g = \langle \gamma \rangle \sqrt{\langle x^2 \rangle \langle x'^2 \rangle - \langle xx' \rangle^2} \quad (1)$$

( $\varepsilon_g$  = rms geometrical emittance) is space charge dominated as long as the space charge collective forces are largely dominant over the emittance pressure, or the laminarity parameter

$$\rho = \frac{\hat{I}}{2I_A \gamma} \frac{\sigma^2}{\varepsilon_n^2} \quad (2)$$

greatly exceeds unity, where  $I_A = 17$  kA is the Alfvén current. Under this condition the beam behaves like a laminar flow (all beam particles move on trajectories that do not cross) and transport and acceleration require a careful tuning of focusing and accelerating elements in order to keep laminarity. Correlated emittance growth is typical in this regime which can be conveniently made reversible [2] if proper beam matching conditions are fulfilled [3]. When  $\rho < 1$  the beam is emittance dominated (thermal regime) and the space charge effects can be neglected. The transition to thermal regime occurs when  $\rho \approx 1$  corresponding to the transition energy

$$\gamma_{tr} = \frac{\hat{I}}{2I_A} \frac{\sigma^2}{\varepsilon_n^2} \quad (3)$$

For example a beam with  $\hat{I} = 100$  A,  $\varepsilon_n = 1 \mu\text{m}$  and  $\sigma = 300 \mu\text{m}$  has a transition energy of 131 MeV. This energy limit defines also the physical extension of the injection system. Space charge effects may recur above transition if bunch compressors are active at higher energies and a new energy threshold with higher  $\hat{I}$  has to be considered.

**DC gun** In a DC gun electrons are emitted by the cathode and are accelerated in the static electric field until they reach an anode with opening to allow beam passage. The limiting current density is given by the modified Child's law [4]

$$J = \frac{4}{9} \varepsilon_0 \sqrt{\frac{2e}{m}} \frac{V^{3/2}}{d^2} F \quad (4)$$

where the enhancement factor  $F \geq 1$  accounts for the finite cathode size. In a high voltage DC gun ( $V = 500$  kV) over a gap of  $d = 5$  cm, the field is about 10 MV/m, giving a limiting current density of 0.33 A/mm<sup>2</sup> with  $F = 1$ . The thermal emittance for a thermionic cathode operating at a temperature  $T$ , is ]

$$\varepsilon_{th} = \frac{R_c}{2} \sqrt{\frac{k_B T}{mc^2}} \quad (5)$$

which gives, per unit cathode size  $R_c$ ,  $\varepsilon_{th} = 0.3 \mu\text{m}/\text{mm}$  at  $T = 2500$  K.

In the *Pierce gun* geometry the cathode has a nonplanar design to provide initial focusing for the beam. DC injectors are characterized by low beam energy extraction and long drift regions used for ballistic bunch compression. For high-current sources there can be significant emittance growth and beam expansion downstream the gun unless care is taken to confine and control the beam with applied magnetic or electrostatic fields.

**Rf gun** In an rf gun electrons are emitted by a photocathode, located inside an rf cavity, typically a one and a half cells standing wave structure, illuminated by a laser pulse so that the bunch length and shape can be controlled on a sub-ps time scale via the laser pulse. The limiting current density is given by (4) or in the short pulse limit at the cathode [15] by

$$J = \frac{e\varepsilon_0 E_0^2}{m_e c} \quad (6)$$

In an S-band rf gun the peak field  $E_0$  is around 100 MV/m, giving an estimated maximum current density in the range of 50 A/mm<sup>2</sup>. The thermal emittance for a metallic photocathode is [5]

$$\varepsilon_{th} = \sigma_x \sqrt{\frac{\hbar\omega - \phi_{eff}}{3mc^2}} \quad (7)$$

where  $\phi_{eff} = \phi_w - \phi_{Schottky}$ ,  $\phi_w$  being the material work function and  $\phi_{Schottky}$  the Schottky work function. With the typical parameters of a Copper photocathode illuminated by a UV laser the expected thermal emittance per unit spot size is around 0.5  $\mu\text{m}/\text{mm}$ .

The emitted electrons are rapidly accelerated to relativistic energies thus partially mitigating the emittance growth due to space charge force effects.

At the gun exit an approximate expression for the space charge induced emittance for a Gaussian bunch of longitudinal and transverse dimensions  $\sigma_z$  and  $\sigma_x$  is [6]

$$\varepsilon_{sc} = \frac{c}{8\alpha\nu_{rf}} \frac{\hat{I}}{I_A} \frac{\sigma_z}{(3\sigma_x + 5\sigma_z)} \quad (8)$$

where  $\alpha = \frac{eE_o}{2mc^2k}$ ,  $E_o$  being the peak rf field,  $k = \frac{c}{2\pi\nu_{rf}}$  with  $\nu_{rf}$  the rf frequency. Notice that  $\alpha \geq 1$  is the typical operational condition of an rf gun.

An important contribution to the beam emittance comes also from time-dependent rf fields which can be evaluated as

$$\varepsilon_x^{rf} = \frac{k^3 \alpha}{\sqrt{2}} \sigma_x^2 \sigma_z^2 \quad (9)$$

Particles are emitted from the cathode with a spread in phase  $\Delta\varphi_o$  that at the gun exit results to be

$$\Delta\varphi_e = \Delta\varphi_o - \frac{\cos\varphi_o}{\alpha \sin^2\varphi_o} \Delta\varphi_o \quad (10)$$

showing that for  $\varphi_o < \pi/2$  (off crest) rf bunch compression occurs.

The induced energy spread  $\sigma_\gamma = \sqrt{\langle\Delta\gamma^2\rangle}$  is

$$\sigma_\gamma = \alpha k \sigma_z \quad (11)$$

with  $\sigma_z = \frac{1}{k} \sqrt{\langle\Delta\varphi_e^2\rangle}$ . The final longitudinal emittance is

$$\varepsilon_z^{rf} = \sqrt{3} (\gamma - 1) k^2 \sigma_z^3 \quad (12)$$

An additional contribution to the transverse emittance comes from the chromatic effects in a solenoid of effective length  $l_{sol}$  and field strength

$k_{sol}^2 = \left(\frac{eB_{sol}}{2\gamma mc}\right)^2$ , usually placed near the gun exit, that in the thin lens approximation scales as

$$\varepsilon_{ch} = \sigma_\gamma k_{sol}^2 l_{sol} \sigma_x^2 \quad (13)$$

Neglecting correlations among all the contributions the total normalized emittance at the gun exit is

$$\varepsilon_n = \sqrt{\varepsilon_{th}^2 + \varepsilon_{sc}^2 + \varepsilon_{rf}^2 + \varepsilon_{ch}^2} \quad (14)$$

**Envelope equation** Under the paraxial ray approximation  $p_x \ll p_z$ , the transverse beam dynamics can be conveniently described by the rms envelope equation that for an axisymmetric beam is [1]

$$\sigma'' + \frac{\gamma'}{\gamma} \sigma' + k_{ext}^2 \sigma = \frac{K_{sc}}{\sigma} + \frac{\varepsilon_g^2}{\sigma^3} \quad (15)$$

The first term is the change in the envelope slope, the second term drives the adiabatic damping due to acceleration, the third term accounts for linear external focusing forces, the fourth represents the defocusing space charge effects and the fifth the internal pressure due to the emittance. In the space charge dominated regime, where  $\rho > 1$ , the emittance term can be neglected. In the transition regime,  $\rho \approx 1$ , both terms should be considered. When  $\rho < 1$  the thermal regime occurs and the space charge term can be neglected.  $K_{sc} = \frac{\hat{I}}{2I_A\gamma^3}$  is the beam perveance,  $\hat{I}$  is the peak current and  $\gamma' = \frac{eE_{acc}}{mc^2}$ ,  $E_{acc}$  being the accelerating field. The rms effective geometrical emittance  $\varepsilon_g^2 = \varepsilon_{th}^2 + \varepsilon_B^2$  includes the contribution of cathode thermal emittance  $\varepsilon_{th}$  and an equivalent emittance term:  $\varepsilon_B = \left(\frac{eB\sigma^2}{8mc}\right)$ , derived from *Busch theorem*, that accounts for the centrifugal potential when the canonical momentum  $p_\theta$ , induced by a residual magnetic field on the cathode is different from zero.

The external focusing  $k_{ext}^2 = k_{sol}^2 + k_{rf}^2$ , in a typical photoinjector design, is due to the solenoid field and to the ponderomotive rf focusing effects described by [7]

$$k_{rf}^2 = \frac{1}{8} \left( \frac{eE_{acc}}{\gamma mc^2} \right)^2 \eta = \frac{1}{8} \left( \frac{\gamma'}{\gamma} \right)^2 \eta$$

where on-crest acceleration only has been considered. The quantity  $\eta$  is a measure of the higher spatial harmonic amplitudes of the rf wave and is generally quite close to unity in standing wave structures and close to 0 in traveling wave structures.

**Brillouin flow** An important case is an intense uniform-density beam matched to an external focusing channel without acceleration ( $k_{ext} = k_{sol}$ ,  $\gamma' = 0$ ,  $\varepsilon_g = \varepsilon_{th}$ ) described by the envelope equation

$$\sigma'' + k_{ext}^2 \sigma = \frac{K_{sc}}{\sigma} + \frac{\varepsilon_{th}^2}{\sigma^3} \quad (16)$$

The Brillouin flow [1] is the equilibrium solution ( $\sigma'' = 0$ ) of (16) given by

$$\sigma = \sqrt{\frac{1}{2} \left( \sigma_{sc}^2 + \sqrt{\sigma_{sc}^4 + 4\sigma_{th}^4} \right)} \quad (17)$$

where

$$\sigma_{sc} = \frac{\sqrt{K_{sc}}}{k_{ext}}, \quad \sigma_{th} = \sqrt{\frac{\varepsilon_{th}}{k_{ext}}} \quad (18)$$

are the two limiting solutions for the space charge and thermal regime respectively. These solutions represent the matching conditions for which the external focusing exactly balances the internal forces.

**Correlated emittance** When transverse-longitudinal correlations within the bunch are important, as the one induced by the space charge and rf fields, a simple model can be considered by slicing the bunch in N longitudinal slices of envelope  $\sigma_s(z, \zeta)$  and divergence  $\sigma'_s(z, \zeta)$  ( $\zeta = \frac{z-z_{tail}}{L}$  is the normalized longitudinal coordinate along a bunch of length L) whose evolution can be computed from a slice envelope equation equivalent to (15) provided that all bunch parameters refer to each individual slice ( $\gamma_s, \gamma'_s, K_{sc,s}$ ). Correlations within the bunch may cause (correlated) emittance oscillations that can be evaluated once an analytical or numerical [8] solution of the slice envelope equation is known by using the definition

$$\varepsilon_{g,cor} = \sqrt{\langle \sigma_s^2 \rangle \langle \sigma'^2_s \rangle - \langle \sigma_s \sigma'_s \rangle^2} \quad (19)$$

where the average is performed over the entire bunch slices ensemble. In the simplest case of a 2 slices model the previous definition reduces to

$$\varepsilon_{g,cor} = |\sigma_1 \sigma'_2 - \sigma_2 \sigma'_1| \quad (20)$$

that represents a very useful formula for an estimation of the emittance scaling.

**Emittance oscillation** An interesting example is the beam dynamics in the space charge dominated regime, nearly matched to an external focusing channel, without acceleration ( $k_{ext} = k_{sol}$ ) [3], described by the slice envelope equation

$$\sigma'' + k_{sol}^2 \sigma_s = \frac{K_{sc,s}}{\sigma_s} \quad (21)$$

A stationary solution is  $\sigma_{s,B} = \frac{1}{k_{sol}^2} \sqrt{\frac{\hat{I}g(\zeta)}{2\gamma^3 I_A}}$  where the local dependence of the current  $\hat{I}_s = \hat{I}g(\zeta)$  within the bunch has been explicitly indicated ( $g(\zeta)$  is the space charge field form factor). Since  $k_{sol}$  has a slice independent constant value, the Brillouin matching condition cannot be achieved at the same time for all the bunch slices. Assuming there is a reference slice perfectly matched with an envelope  $\sigma_{r,B}$ , the matching condition for the other slices can be written as  $\sigma_{sB} = \sigma_{rB} + \frac{\sigma_{rB}}{2} \left( \frac{\delta I_s}{I} \right)$  with respect to the reference slice. Considering a small perturbation  $\delta_s$  from the equilibrium in the form  $\sigma_s = \sigma_{s,B} + \delta_s$

and substituting in Eq.(21) we obtain an equation for the slice offset:  $\delta_s'' + 2k_{ext}^2 \delta_s = 0$ . The solution  $\delta_s = \delta_o \cos(\sqrt{2}k_{ext}z)$ , where  $\delta_o = \sigma_{so} - \sigma_{sB}$  is the amplitude of the initial mismatch that we assume for simplicity the same for all slices, leads to the expression

$$\sigma_s = \sigma_{s,B} + \delta_o \cos(\sqrt{2}k_{ext}z) \quad (22)$$

Equation(22) shows that slice envelopes oscillate around the equilibrium solution with the same (plasma) wavelength for all slices which depends only on the external focusing forces. Using the two slices model and substituting solution (22), with  $s = 1, 2$ , in Eq.(20) the emittance evolution results,

$$\varepsilon_{g,cor} = \frac{1}{4} k_{sol} \sigma_{rB} \left| \frac{\Delta I}{\hat{I}} \delta_o \sin(\sqrt{2}k_{ext}z) \right| \quad (23)$$

where  $\Delta I = \hat{I}_1 - \hat{I}_2$ . Note that in this simple case envelope oscillations of the mismatched slices induce undamped correlated emittance oscillations which periodically goes back to zero. Emittance oscillations are the basic features of the emittance compensation technique.

**Emittance compensation** Space charge induced emittance growth is partially correlated and can be reduced in a drift downstream the electron source with a solenoid located at the exit of the gun [2]. In order to prevent additional space charge emittance growth in the subsequent accelerating sections, emittance oscillations have to be damped and the final emittance minimum has to be reached at the transition energy so that space charge forces become ineffective. To this end the beam has to be properly matched to the accelerating sections (Brillouin like flow) in order to keep under control emittance oscillations via the ponderomotive rf focusing force [3]. In many cases rf focusing is too weak to provide sufficient beam containment. A long solenoid around the accelerating structure is a convenient replacement to provide the necessary focusing. This possibility is included in the following discussion.

The matching conditions for the beam subject to acceleration ( $\gamma = \gamma_0 + \gamma' z$ ) can be derived following the previous example. With the substitution of the reduced variable  $\hat{\sigma} = \sigma \sqrt{\gamma}$  [9], the envelope equation (15) becomes

$$\hat{\sigma}'' + \hat{k}_{ext}^2 \hat{\sigma} = \frac{\hat{K}_{sc}}{\hat{\sigma}} + \frac{\varepsilon_n^2}{\hat{\sigma}^3} \quad (24)$$

which is formally identical to Eq.(15) with the scaled parameters  $\hat{k}_{ext}^2 = k_{ext}^2 + \frac{1}{4} \left( \frac{\gamma'}{\gamma} \right)^2 = k_{sol}^2 + \left( \frac{\gamma'}{\gamma} \right)^2 \left( \frac{1}{4} + \frac{\eta}{8} \right)$ ,  $\varepsilon_n = \gamma \varepsilon_g$  and  $\hat{K}_{sc} = \gamma K_{sc}$ . Equilibrium solutions in the reduced variables (“invariant envelope” in the literature [3]) are  $\hat{\sigma}_{sc} = \frac{\sqrt{\hat{K}_{sc}}}{\hat{k}_{ext}}$  and  $\hat{\sigma}_{th} = \sqrt{\frac{\varepsilon_n}{\hat{k}_{ext}}}$  in the space charge and thermal regime respectively, corresponding to the matching conditions for the beam envelope

$$\sigma_{sc} = \sqrt{\frac{2\hat{I}/\gamma I_A}{(\Theta^2 + \gamma'^2 (\frac{\eta}{2} + 1))}} \text{ for } \rho > 1 \quad (25)$$

where  $\Theta = \frac{eB}{mc}$  and

$$\sigma_{th} = \sqrt{\frac{\varepsilon_n}{(\Theta^2 + \gamma'^2 (\frac{\eta}{2} + 1))^{1/2}}} \text{ for } \rho < 1 \quad (26)$$

Note that condition (25) scales like  $\gamma^{-1/2}$  while (26) is independent of  $\gamma$ . For the transition regime ( $\rho \approx 1$ ) the matching condition can be derived from an equivalent expression of (17) with reduced variables.

An approximate expression for emittance oscillations in the space charge dominated regime, i.e. when  $\gamma < \gamma_{tr}$ , can be derived in a similar way as (23) taking into account acceleration [3], giving

$$\Delta \varepsilon_n(z) \propto \frac{\delta_0}{\gamma'} \sqrt{\frac{I/I_A}{2\gamma}} \left| \cos \psi - \sqrt{2} \sin \psi \right| \quad (27)$$

where  $\psi \equiv \ln(\gamma/\gamma_0)/\sqrt{2}$ . Before the transition energy is achieved the emittance performs damped oscillations with wavelength depending on the external fields and with amplitude depending on the current profile. A careful tuning of the external fields and bunch charge profile can minimize the value of the emittance at the injector extraction. A successful application of the emittance compensation technique can be seen in [10, 11].

The previous description of emittance compensation for space charge effects applies equally well to both rf and DC guns.

**Velocity bunching** (VB) Emittance compensation is a key feature of the Velocity Bunching technique [12, 13], a method to compress the bunch using rectilinear trajectories in an rf structure by means of the longitudinal rf focusing effect. Under the assumption that the beam current

Table 1: Rf and wavelength scaling laws.

|                                | Charge            | Wavelength             |
|--------------------------------|-------------------|------------------------|
| $Q$                            | $\propto Q$       | $\propto \lambda$      |
| $\sigma_x, \sigma_y, \sigma_z$ | $\propto Q^{1/3}$ | $\propto \lambda$      |
| $E_o$                          |                   | $\propto \lambda^{-1}$ |
| $B$                            |                   | $\propto \lambda^{-1}$ |
| $\varepsilon_{sc}$             | $\propto Q^{2/3}$ | $\propto \lambda$      |
| $\varepsilon_{th}$             | $\propto Q^{1/3}$ | $\propto \lambda$      |
| $\varepsilon_{rf}$             | $\propto Q^{4/3}$ | $\propto \lambda$      |

grows as a linear function of the energy like  $\hat{I} = \frac{\gamma}{\gamma_o} \hat{I}_o$  the matching condition in the space charge dominated regime is given by

$$\sigma_{sc} = \sqrt{\frac{2\hat{I}_o}{\gamma_o I_A (\Theta^2 + \gamma'^2 (\frac{\eta}{2} + 1))}} \quad (28)$$

where  $\gamma_o$  and  $I_o$  are the values for the current and the energy respectively at injection into the VB structure. Also this solution is independent of  $\gamma$ .

**Charge and wavelength scaling laws** Once an optimized rf injector design has been found, a simple procedure allows to keep the optimized performances with a different bunch charge  $Q$  and/or rf wavelength  $\lambda$ , by simply scaling the relevant parameters as shown in Tab.1 [14]. The resulting emittance scaling is also reported.

## References

- [1] M. Reiser, Theory and Design of Charged Particle Beams, Wiley, New York, 1994
- [2] B.E. Carlsten, NIM A285, 313 (1989)
- [3] L. Serafini, J.B. Rosenzweig, PR E55, 7565 (1997)
- [4] K. Togawa et al, PRST-AB 10, 020703 (2007)
- [5] D.H. Dowell et al, PRST-AB 12, 074201 (2009)
- [6] K.J. Kim, NIM A275, 201 (1989)
- [7] J.B. Rosenzweig, L. Serafini, PR E49, 1599 (1994)
- [8] M. Ferrario et al, Int. Journal of Modern Physics A, Vol. 22, No. 3 (2007)
- [9] C. Wang, PRE E4, 046502 (2006)
- [10] R. Akre et al, PRST-AB 11, 030703 (2008)
- [11] M. Ferrario et al, PRL 99, 234801 (2007)
- [12] L. Serafini, M. Ferrario, AIP Conf. Proc. 581, 87 (2001)
- [13] M. Ferrario et al, PRL 104, 054801 (2010)
- [14] J.B. Rosenzweig, E. Colby, report TESLA-95-04
- [15] J.B. Rosenzweig et al, NIM A57, 87 (2006)

## 2.4.5 Space Charge Effects in Circular Accelerators

B. Zotter, CERN

### 2.4.5.1 Direct space charge effects

**Impedances** The electric and magnetic forces ( $F_E$  and  $F_M = -\beta^2 F_E$ ) due to a straight beam of charged particles nearly cancel when their velocity  $v = \beta c$  is ultrarelativistic ( $\beta \approx 1$ ), both in free space and in a perfectly conducting, concentric smooth pipe. The factor  $1 - \beta^2 = 1/\gamma^2 \ll 1$  then reduces the *direct space-charge force* and hence the longitudinal space-charge impedance (See also Sec.3.2.4)

$$\frac{Z_{||}}{n} = \frac{i Z_0}{2\beta\gamma^2} g \quad (1)$$

where  $n = \omega/\omega_0$ . The *g-factor* is  $g = 1 + 2\ln(b/a)$  for a circular cylindrical beam of radius  $a$  in a concentric pipe of radius  $b > a$ .

Space-charge force is quite weak in proton machines with GeV energies, and completely negligible in high-energy electron machines. However, in proton or ion machines in the MeV range, space charge is often the dominant impedance. It corresponds to a negative inductance, sometimes called *capacitive* although it is  $\propto \omega$  like an inductance. In machines operating above transition this may lead to *negative mass instability* (Sec.2.4.9).

The transverse space-charge impedance is

$$Z_{\perp} = i \frac{R Z_0}{\beta^2 \gamma^2} \left( \frac{1}{a^2} - \frac{1}{b^2} \right) \quad (2)$$

where  $R$  = average machine radius. For a uniform beam with elliptic cross section [1], half axes  $a_1$  and  $a_2$ , the first term in brackets should be replaced by  $2/a_{1,2}(a_1 + a_2)$ , and by  $1/\sigma_{x,y}(\sigma_x + \sigma_y)$  for a Gaussian distribution.

Betatron tune shifts are proportional to  $Z_{\perp} I_b$  and can bring tunes onto resonances leading to particle losses. In high-current machines, the tune variation can be reduced by surrounding the beam with an *rf shield* following the beam profile such that the ratio  $b/a$  remains constant. In bunched beams, tunes are different in the center and at the edge of a bunch, and the whole tune region should avoid low order resonances.

The direct space charge force is centered on the beam and not the chamber axis. Hence it is independent of any transverse displacement of the beam and does not affect rigid dipole oscillations. However, it changes the external focusing forces

which leads to the *beam envelope equation* for the beam half-width  $a_x$  (see also Sec.2.4.4),

$$\frac{d^2 a_x}{ds^2} + K_x(s) a_x - \frac{\epsilon_x^2}{a_x^3} - \frac{4\lambda r_0}{\beta^2 \gamma^3 (a_x + a_y)} = 0 \quad (3)$$

The equation for the beam half-height  $a_y$  is obtained by interchanging all subscripts  $x$  and  $y$ . Here  $K_i$  is the external focusing strength, including possible gradient errors,  $\epsilon_{x,y}$  the emittances in the  $x$  or  $y$  plane,  $\lambda$  the (constant) line density,  $r_0$  the classical particle radius, and  $\gamma$  the energy factor. The space charge force leads to an *envelope modulation* which reduces the tune shift to  $\frac{3}{4}\Delta\nu_{sc}^{(0)}$  for the antisymmetric (“quadrupole”) mode, and to  $\frac{1}{2}\Delta\nu_{sc}^{(0)}$  for the symmetric (“breathing”) mode, where  $\Delta\nu_{sc}^{(0)}$  is the space-charge tune shift when the modulation is neglected.

**Curvature effects** The compensation of electric and magnetic forces is perturbed when the beam trajectory and the surrounding vacuum chamber are curved. The impedance of a perfectly conducting chamber of rectangular cross section (half height  $h$ , half width  $w$ ) bent in a circle of radius  $R$  is purely reactive [2],

$$\text{Im} \frac{Z_{\parallel}}{n} \approx Z_0 \left( \frac{2h}{\pi R} \right)^2 \left[ A - 3B \left( \frac{2h\omega}{\pi c} \right)^2 \right] \quad (4)$$

where the constants  $A$  and  $B$  are close to unity when  $w \geq 2h$ , and about 0.7 when  $w = h$ . Eq.(4) holds for  $k = \omega/\beta c < (\pi/h)\sqrt{R/w}$ , corresponding typically to a frequency of 100 GHz. At higher frequencies, there are resonances with  $R/Q$  values given in terms of Bessel function zeros of large order [3]. Based on a simpler model of a circular beam between two infinite plates separated by  $2g$ , the real part of the *radiation impedance* can be approximated by [4]

$$\text{Re} \frac{Z_{\parallel}}{n} = 300 [\Omega] \frac{g}{R} \quad (5)$$

The effect of curvature on transverse beam motion due to the *centrifugal space charge force* was first considered important [5], but it is cancelled almost completely by that of the transverse electric field on particles with different energies [6]. Nevertheless, curved trajectories can lead to emittance growth due to additional forces [7]. The energy gradient along a Gaussian bunch becomes

$$\frac{dE}{cdt} = \frac{2N_b E_0 r_0}{\sqrt{2\pi}(3R^2\sigma_s^4)^{1/3}} F_0 \left( \frac{s}{\sigma_s} \right) \quad (6)$$

while the additional transverse forces are

$$F_x(s) = -\frac{N_b E_0 r_0}{\sqrt{2\pi} R \sigma_s} \left[ 2e^{-\frac{s^2}{2\sigma_s^2}} + \frac{3x F_1(s/\sigma_s)}{(9R\sigma_s^2)^{1/3}} \right] \\ F_y(s) = -\frac{N_b E_0 r_0}{\sqrt{2\pi} R \sigma_s} \frac{y}{(9R\sigma_s^2)^{1/3}} F_1 \left( \frac{s}{\sigma_s} \right) \quad (7)$$

where

$$F_0(v) = \int_0^\infty \frac{du}{u^{1/3}} (v-u) e^{-(v-u)^2} \\ F_1(v) = \int_0^\infty \frac{du}{u^{2/3}} e^{-(v-u)^2} \quad (8)$$

These transverse forces as well as the longitudinal energy gradient in dispersive regions contribute to emittance growth.

#### 2.4.5.2 Betatron frequency shifts

In addition to direct space-charge fields, *induced fields* due to charges and currents in the surrounding vacuum chamber walls (“image fields”) shift the betatron frequencies, which is often called *Laslett tune shifts* [8]. Below we present the vertical tune shifts; the horizontal ones can be obtained by exchanging  $y \leftrightarrow x$  and height  $\leftrightarrow$  width.

**Incoherent tune shifts** refer to changes of betatron frequencies of single particles in a beam. Let  $\omega_y = \nu_y \omega_0$  be the unperturbed betatron frequency,  $F_y$  the vertical forces due to induced and space-charge fields,  $I_0$  the beam current, and  $y_0$  the offset of beam centroid. To first order,

$$F_y = y \frac{\partial F_y}{\partial y} \Big|_{y=y_0=0} + y_0 \frac{\partial F_y}{\partial y_0} \Big|_{y=y_0} \quad (9)$$

The first term gives rise to a change of the *incoherent betatron tune*  $\nu_{\text{inc}}$  with

$$\nu_{\text{inc}}^2 = \nu_y^2 - \frac{1}{m_0 \gamma \omega_0^2} \frac{\partial F_y}{\partial y} \Big|_{y=y_0=0} \quad (10)$$

When the shift  $\Delta\nu_{\text{inc}} \ll \nu_y$  is small,  $\Delta\nu_{\text{inc}} \approx (\nu_y^2 - \nu_{\text{inc}}^2)/2\nu_{\text{inc}}$ . Incoherent tunes of different particles in a beam usually vary over a finite range called *tune spread*.

**Coherent tune shifts** change the frequency of the beam centroid. Measurements of tunes, e.g. with pick-up electrodes sensitive only to average motion, show the coherent tune which satisfies

$$\nu_{\text{coh}}^2 = \nu_y^2 - \frac{1}{m_0 \gamma \omega_0^2} \left[ \frac{\partial F_y}{\partial y} \Big|_{y=y_0=0} + \frac{\partial F_y}{\partial y_0} \Big|_{y=y_0} \right] \quad (11)$$

For small shifts,  $\Delta\nu_{\text{coh}} \approx (\nu_y^2 - \nu_{\text{coh}}^2)/2\nu_{\text{coh}}$ .

The difference between  $\omega_{\text{inc}}$  and  $\omega_{\text{coh}}$  is proportional to  $\partial F_y/\partial y$  and  $\partial F_y/\partial y_0$  and equals

Table 1: Laslett coefficients for simple geometries.

| Coefficient          | circular | elliptic cross section  | par.plate      | comments            |
|----------------------|----------|---|----------------|---------------------|
| $\varepsilon_{sc,V}$ | 1/2      | $b/(a+b)$   | -              | direct space charge |
| $\varepsilon_{sc,H}$ | 1/2      | $b^2/a(a+b)$  | -              |                     |
| $\varepsilon_{1V/H}$ | 0        | $\pm \frac{h^2}{12d^2} \left[ (1+k'^2) \left( \frac{2K}{\pi} \right)^2 - 2 \right]$ | $\pm \pi^2/48$ | incoherent electric |
| $\varepsilon_{2V/H}$ | -        | -   | $\pm \pi^2/24$ | incoherent magnetic |
| $\xi_{1V}$           | 1/2      | $\frac{h^2}{4d^2} \left[ \left( \frac{2K}{\pi} \right)^2 - 1 \right]$               | $\pi^2/16$     | coherent electric   |
| $\xi_{1H}$           | 1/2      | $\frac{h^2}{4d^2} \left[ 1 - \left( \frac{2Kk'}{\pi} \right)^2 \right]$             | 0              |                     |
| $\xi_{2V}$           | -        | -   | $\pi^2/16$     | coherent magnetic   |
| $\xi_{2H}$           | -        | -   | 0              |                     |

$U + (1+i)V$ , where  $U$  and  $V$  are the *transverse dispersion relation coefficients*, proportional to the transverse impedance  $Z_\perp$ ,

$$U + (1+i)V = i \frac{Z_\perp}{Z_0} \frac{r_0 I_0}{\gamma \nu_y e} \quad (12)$$

**Unbunched beams** The vertical tune shifts for unbunched beams in a perfectly conducting vacuum chamber of half-height  $h$ , between perfect magnetic pole pieces at a distance  $\pm g$  from the axis, can be written

$$\Delta\nu_{\text{inc}} = -\frac{2r_0}{ec} \frac{RI_0 \langle \beta \rangle}{\beta^3 \gamma} \left[ \frac{\varepsilon_1}{h^2} + \beta^2 \frac{\varepsilon_2}{g^2} + \frac{\varepsilon_{sc}}{\gamma^2 b^2} \right] \quad (13)$$

$$\Delta\nu_{\text{coh}} = -\frac{2r_0}{ec} \frac{RI_0 \langle \beta \rangle}{\beta^3 \gamma} \left\{ \begin{aligned} & \frac{\xi_1}{h^2} + \frac{\beta^2 \xi_2}{g^2} \\ & \frac{\beta^2 \varepsilon_1}{h^2} + \frac{\xi_1}{\gamma^2 h^2} + \frac{\beta^2 \varepsilon_2}{g^2} \end{aligned} \right\}$$

where  $\langle \beta \rangle$  is the average  $\beta$ -function (=  $R/\nu$  in smooth approximation),  $r_0$  the classical radius of the particle. The first expression for  $\Delta\nu_{\text{coh}}$ , called *integer formula*, is valid for very thin walls through which ac magnetic fields can penetrate, while the second one, called *half-integer formula*, is for the usual case of thicker walls where these fields do not penetrate [9]. The transition is discussed in [9].

The *Laslett coefficients* for beams in the center of some common geometries are shown in Tab.1, where  $K(k)$  is the complete elliptic integral of the first kind with modulus  $k$ , given by  $\exp(-\pi K'/K) = (w-h)/(w+h)$ , where  $h$

and  $w$  are the chamber half height and width,  $k' = \sqrt{1-k^2}$  is the *complementary modulus*, and  $K' = K(k')$ . [This relation is much simpler when expressed with the *nome*, defined as  $q = \exp(-\pi K'/K)$ , which is a common parameter in most computer codes for elliptic integrals.]

When a beam is partially neutralized, and  $\eta$  expresses the ratio of densities of particles of opposite charge to particles in the beam, the electric fields due to direct space-charge and due to induced charges are reduced, while the magnetic field remains unchanged. This can strongly increase the direct space-charge field, which is no longer multiplied by  $1 - \beta^2 = 1/\gamma^2$  but by  $1 - \beta^2 - \eta$ . To avoid excessive tune shifts when  $\eta \gg 1/\gamma^2$  one needs to reduce neutralization, e.g. with clearing electrodes or lower rest gas pressure. Tune shifts of unbunched beams with neutralization are given in Eq.(14) with  $B = 1$ . Image coefficients for off-axis beams can be found in [10, 11].

The image coefficients  $\varepsilon_2$  and  $\xi_2$  for closed magnetic boundaries (e.g. circular or elliptic) cannot be calculated for  $\mu \rightarrow \infty$  since the induced magnetic field would not permit a charged beam to pass through. They have therefore been calculated for perfect magnetic yokes with gaps [12]. Closed magnetic yokes are used e.g. in superconducting magnets, but there the coefficients  $\varepsilon_2 = \xi_2 \rightarrow 0$  since the magnetic material is driven completely into saturation ( $\mu \rightarrow 1$ ).

**Bunched beams** For bunched beams, several terms in the expression for the tune shifts are proportional to the peak rather than the average current. This can be taken into account by division with the bunching factor  $B < 1$ , defined as the ratio of average to peak current. This yields

$$\Delta\nu_{\text{inc}} = -\frac{2r_0}{ec} \frac{RI_0\langle\beta\rangle}{\beta^3\gamma} \left[ \left( \beta^2 + \frac{1-\beta^2-\eta}{B} \right) \frac{\varepsilon_1}{h^2} + \beta^2 \frac{\varepsilon_2}{g^2} + \frac{1-\beta^2-\eta}{B} \frac{\varepsilon_{sc}}{b^2} \right] \quad (14)$$

$$\Delta\nu_{\text{coh}} = -\frac{2r_0}{ec} \frac{RI_0\langle\beta\rangle}{\beta^3\gamma} \times \begin{cases} \left( \beta^2 + \frac{1-\beta^2-\eta}{B} \right) \frac{\xi_1}{h^2} + \beta^2 \frac{\xi_2}{g^2} \\ \beta^2 \frac{\varepsilon_1}{h^2} + \frac{1-\beta^2-\eta}{B} \frac{\xi_1}{h^2} + \beta^2 \frac{\varepsilon_2}{g^2} \end{cases}$$

The neutralization in bunched beams is usually weaker than in unbunched ones, and can be further reduced by leaving a *gap* in the filling pattern.

For machines with vacuum chambers consisting of sections with unequal dimensions, one needs to calculate the tune shifts for each chamber type separately, and add the contributions in proportion to their lengths.

## References

- [1] L. Teng, Report ANLAD-59 (1963)
- [2] Y.K. Ng, R. Warnock, PRD 40 (1989) 231
- [3] R. Warnock, P. Morton, SLAC PUB 4562 (1988)
- [4] A. Faltens, L.J. Laslett, PA 4 (1973) 152
- [5] R. Talman, PRL 56 (1986) 1429
- [6] E. Lee, PA 25 (1990) 241
- [7] V. Derbenev, Y. Shiltsev, SLAC-PUB 7181(1996); P. Emma, R. Brinkman, PAC 97; R. Li, C.L. Bohn, J.I. Bisognano, PAC 97
- [8] L.J. Laslett, BNL Report 7534 (1963) p.324
- [9] B. Zotter, CERN/ISR-TH/72-8 (1972)
- [10] B. Zotter, CERN/ISR-TH/74-11 and 38 (1972)
- [11] G. Guignard, CERN Report 77-10 (1977)
- [12] L.J. Laslett, Report UCRL 18892 (1969)

### 2.4.6 Beam Dynamics in Proton Linacs

S. Henderson, FNAL  
A. Aleksandrov, ORNL

Proton linacs differ from electron linacs in several ways, all of which are due to the fact that the proton beam is not ultrarelativistic: i) the spacing between accelerating gaps increases with particle velocity, ii) space-charge effects strongly

influence the beam dynamics, iii) longitudinal phase-stability is maintained by accelerating off-crest with negative synchronous phases, and iv) radial defocusing in accelerating gaps must be taken into account [1, 2, 3, 4].

**Energy gain and transit time factor** For excitation of an accelerating gap (considered a cylindrical pillbox cavity) in a TM<sub>010</sub> standing wave mode, the non-zero fields are written  $E_z(r, z) \cos(\omega t + \phi)$ ,  $E_r(r, z) \cos(\omega t + \phi)$  and  $B_\theta(r, z) \sin(\omega t + \phi)$ . Assume an accelerating gap of length  $L$  with the origin  $z = 0$  at the electrical center of the cell. Consider a particle traversing the accelerating gap such that at  $t = 0$  it is located at  $z = 0$ . The phase of the rf field is  $\phi$ . The energy gain in the gap is

$$\Delta W = qV_0 T(r, k) \cos \phi$$

where  $k = \omega/c\beta = 2\pi/\beta\lambda$ ,

$$V_0 = E_0 L = \int_{-L/2}^{L/2} E_z(0, z) dz$$

$E_0$  is the peak axial field, and  $T(r, k)$  is the *transit-time factor*,

$$T(r, k) = \frac{1}{V_0} \int_{-L/2}^{L/2} E_z(r, z) \cos kz dz$$

With cylindrical symmetry,

$$T(r, k) = T(0, k) I_0(Kr) = T(k) I_0(Kr)$$

where  $T(k)$  is the *axial transit-time factor*, and

$$K^2 = k^2 - k_0^2 = \left( \frac{2\pi}{\gamma\beta\lambda} \right)^2$$

where  $k_0 = \omega/c$ .

For  $E_z$  constant within the accelerating gap at the drift tube bore radius  $r = a$  [ $E_z(a, |z| < L/2) = \text{const}$ ] but zero outside the gap,

$$T(r, k) = J_0 \left( \frac{2\pi a}{\lambda} \right) \frac{I_0(Kr)}{I_0(Ka)} \frac{\sin(\pi L/\beta\lambda)}{\pi L/\beta\lambda} \quad (1)$$

which gives for the axial transit-time factor, in the limit  $a/\lambda \rightarrow 0$ , the familiar result

$$T(k) = \frac{\sin(\pi L/\beta\lambda)}{\pi L/\beta\lambda}$$

When considering multiple accelerating gaps, it is useful to define the *synchronous particle* as that ideal particle which arrives at each gap with the design beam/rf relative phase (known as the *synchronous phase*) prescribed in the “phase law,” and which therefore acquires the design energy gain in each gap. The synchronous particle has phase and energy coordinates  $\phi_s$  and  $W_s$  and velocity  $c\beta_s$ .

The transit time factor for a multicell structure is given by Eq.(1) with  $K$  replaced by  $K_s = 2\pi/\gamma_s \beta_s \lambda$ .

The transit time factor for a multicell  $\pi$ -mode structure (e.g. elliptical multi-cell cavity) with field  $E_z(z) = E_0 \cos k_s z$ , where  $k_s = \pi/L$ , is

$$T(r, k) = \frac{\pi}{4} I_0(K_s r) \left( \frac{\sin((\beta_s/\beta - 1)\pi/2)}{(\beta_s/\beta - 1)\pi/2} + \frac{\sin((\beta_s/\beta + 1)\pi/2)}{(\beta_s/\beta + 1)\pi/2} \right) \quad (2)$$

which gives, for  $\beta = \beta_s$ ,  $T(0, k_s) = \pi/4$ .

### Longitudinal dynamics (See also Sec.2.1.2)

Consider the motion of a particle with phase and energy coordinates  $\phi(s)$  and  $W(s)$ . Define conjugate particle coordinates relative to those of the synchronous particle:  $\varphi(s) = \phi(s) - \phi_s(s)$  and  $w(s) = W(s) - W_s(s)$ . In smooth approximation with small acceleration rate, and assuming that  $E_0 T$ ,  $\phi_s$ , and  $\beta_s$  are constant, the equations of motion become

$$\varphi'(s) = -\frac{2\pi}{mc^2 \beta_s^3 \gamma_s^3 \lambda} w = \frac{\partial H}{\partial w}$$

$$w'(s) = -qE_0 T [\cos \phi_s (1 - \cos \varphi) + \sin \phi_s \sin \varphi] = -\frac{\partial H}{\partial \varphi}$$

which derive from a Hamiltonian

$$H = -\frac{2\pi}{mc^2 \beta_s^3 \gamma_s^3 \lambda} \frac{w^2}{2} - qE_0 T [\sin \phi_s (\cos \varphi - 1) + \cos \phi_s (\sin \varphi - \varphi)] \quad (3)$$

Fig.1 shows the separatrix and resulting phase space trajectories, which display stable motion about the synchronous phase  $\phi_s$ . The phase acceptance of the separatrix extends from  $\varphi_{\max} = \phi_{\max} - \phi_s = -2\phi_s$  to  $\varphi_{\min} = \phi_{\min} - \phi_s$ , which is given by solution of the equation

$$\sin \phi_{\min} - \phi_{\min} \cos \phi_s + \sin \phi_s - \phi_s \cos \phi_s = 0$$

The total phase acceptance,  $\Psi = \varphi_{\max} - \varphi_{\min}$ , is related to the design synchronous phase by

$$\tan \phi_s = \frac{\sin \Psi - \Psi}{1 - \cos \Psi}$$

which for  $\Psi \ll 1$  is  $\tan \phi_s \approx -\Psi/3$ . Note that the phase acceptance shrinks to zero as  $\phi_s \rightarrow 0$ .

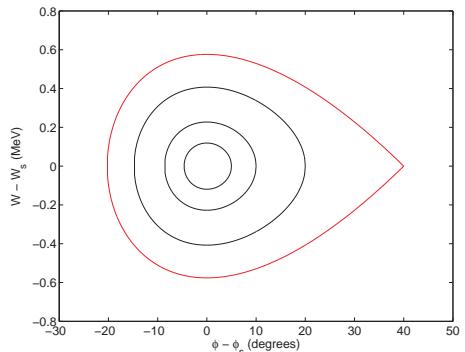


Figure 1: Longitudinal phase space trajectories and separatrix (red) with  $\phi_s = -20^\circ$ .

The half-height of the separatrix at  $\phi_s$ , the *energy acceptance*, is

$$w_{\max}^2 = \frac{2qE_0 T mc^2 \beta_s^3 \gamma_s^3 \lambda}{\pi} (\phi_s \cos \phi_s - \sin \phi_s)$$

For small amplitude oscillation, the equation of motion for the phase becomes  $\varphi'' + k_{z0}^2 \varphi = 0$  with the *longitudinal zero-current phase advance per meter*

$$k_{z0}^2 = \frac{2\pi qE_0 T \sin(-\phi_s)}{mc^2 \beta_s^3 \gamma_s^3 \lambda}$$

The phase advance per focusing period of length  $l$  is  $\nu_{z0} = k_{z0} l$ . The phase-space trajectory is an ellipse of the form  $w^2/\hat{w}^2 + \varphi^2/\dot{\varphi}^2 = 1$  with  $\hat{w}^2 = qE_0 T mc^2 \beta_s^3 \gamma_s^3 \lambda \sin(-\phi_s) \dot{\varphi}^2 / 2\pi$ . The area of the phase space ellipse is an adiabatic invariant, so that after acceleration, an ellipse with initial phase and energy widths  $\varphi_i, \hat{w}_i$  is transformed to an ellipse with final widths given by  $\hat{\varphi}_f/\dot{\varphi}_i = (\beta_i \gamma_i / \beta_f \gamma_f)^{3/4}$  and  $\hat{w}_f/\hat{w}_i = (\beta_f \gamma_f / \beta_i \gamma_i)^{3/4}$ . The phase motion is damped at the expense of an increase in energy spread.

When acceleration is included, that is, when  $\beta \gamma \neq \text{const}$  in Eq.(3), the shape of the input acceptance resembles a golf-club, as shown in Fig.2. Also shown in Fig.2 are two trajectories which demonstrate the phase damping.

**Transverse beam dynamics** Transverse defocusing terms arise from the  $E_r$  and  $B_\theta$  rf field components in the accelerating gaps. The radial defocusing is given by

$$\Delta(\gamma \beta_r) = -\frac{qE_0 T(k) L I_1(Kr) \sin \phi}{mc^2 \beta_s \gamma_s}$$

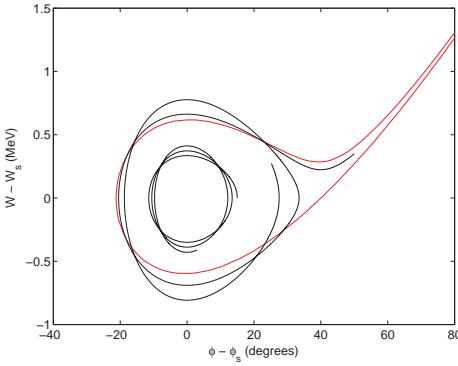


Figure 2: Longitudinal phase space trajectories and input acceptance (red) for adiabatic acceleration with  $\phi_s = -20^\circ$ .

which for  $Kr \ll 1$  gives

$$\Delta(\gamma\beta_r) = -\frac{\pi qE_0T(k)L \sin \phi}{mc^2\beta_s^2\gamma_s^2\lambda} r$$

The rf defocusing is more severe at low energies, and must be taken into account in the linac design. The combined focusing from quadrupoles or solenoids and rf gaps is included in determining transverse stability of the motion via stability diagrams [5, 6].

**Gap transformation equations** Proton dynamics in an accelerating gap may be calculated using a thin-lens method in which the beam is assumed to drift at constant velocity to the accelerating gap center, where the energy, phase and radial coordinates are advanced, followed by drift at the new velocity to the end of the cell [7, 3, 8]. The initial coordinates at the entrance to the accelerating cell are  $W_i, \phi_i, r_i, r'_i$ , the coordinates at the exit of the cell are  $W_e, \phi_e, r_e, r'_e$ , and those at the gap center are  $W_c, \phi_c, r_c, r'_c$ . The gap transformations are

$$\begin{aligned} W_e &= W_i + qV_0 T(r_c, k_c) \cos \phi_c \\ \phi_e &= \phi_i + \frac{L}{2} \frac{2\pi}{\lambda} \left( \frac{1}{\beta_i} + \frac{1}{\beta_e} \right) \\ &\quad + \frac{qV_0}{mc^2\beta_c^2\gamma_c^3} k_c T'(r_c, k_c) \sin \phi_c \\ r_e &= r_i + \frac{L}{2} (r'_i + r'_e) - \frac{qV_0}{mc^2\beta_c^2\gamma_c^3} T'_r(r_c, k_c) \cos \phi_c \\ r'_e &= r'_i - \frac{qV_0}{mc^2\beta_c^2\gamma_c^3} T_r(r_c, k_c) \sin \phi_c \end{aligned} \quad (4)$$

and

$$\begin{aligned} W_c &= W_i + \frac{qV_0}{2} [T(r_c, k_c) \cos \phi_c \\ &\quad + S(r_c, k_c) \sin \phi_c] \\ \phi_c &= \phi_i + \frac{L}{2} \frac{2\pi}{\beta_i \lambda} + \frac{qV_0 k_c}{2mc^2\beta_c^2\gamma_c^3} [T'(r_c, k_c) \sin \phi_c \\ &\quad - S'(r_c, k_c) \cos \phi_c] \\ r_c &= r_i + \frac{L}{2} r'_i - \frac{qV_0}{2mc^2\beta_c^2\gamma_c^3} [T'_r(r_c, k_c) \cos \phi_c \\ &\quad - S'_r(r_c, k_c) \sin \phi_c] \\ r'_c &= r'_i - \frac{qV_0}{2mc^2\beta_c^2\gamma_c^3} [T_r(r_c, k_c) \sin \phi_c \\ &\quad + S_r(r_c, k_c) \cos \phi_c] \end{aligned} \quad (5)$$

where

$$\begin{aligned} S(r, k) &= \frac{2}{V_0} \int_0^{L/2} dz E_z(r, z) \sin kz \\ T_r(r, k) &= \frac{2}{V_0} \int_0^{L/2} dz E_r(r, z) \sin kz \\ S_r(r, k) &= \frac{2}{V_0} \int_0^{L/2} dz E_r(r, z) \cos kz \end{aligned} \quad (6)$$

and a prime denotes taking derivative with respect to  $k$ . For cylindrical symmetry,  $T_r(r, k) = T(k)kI_1(Kr)/K$ . The full and mid-gap transformation equations are coupled and may be solved by iteration. In the above transformations, (small) gap correction terms proportional to  $r'$  have been neglected. See [7, 3] for the full treatment.

**Space charge dynamics** The effect of space charge forces on beam dynamics decreases proportional to  $1/\beta^2\gamma^3$  and is therefore important for nonrelativistic beams in high-intensity accelerators.

The Kapchinskii-Vladimirskiy (K-V) envelope equations [9] describe horizontal,  $X$ , and vertical,  $Y$ , beamsizes (ellipse semi-axes) of a continuous uniform density beam in a periodic quadrupole-focusing channel,

$$\begin{aligned} \frac{d^2X}{ds^2} + k_x(s)X - \frac{\epsilon_x^2}{X^3} - \frac{2K}{X+Y} &= 0 \\ \frac{d^2Y}{ds^2} + k_y(s)Y - \frac{\epsilon_y^2}{Y^3} - \frac{2K}{X+Y} &= 0 \end{aligned} \quad (7)$$

where  $k_{x,y}(s)$  represents the external focusing force,  $\epsilon_{x,y}$  is the total emittance, and

$$K = \frac{qI}{2\pi\epsilon_0 mc^3} \frac{1}{\beta^3\gamma^3}$$

is the *generalized permeance*.

The K-V envelope equations can be generalized to an arbitrary charge density distribution with *elliptical symmetry* in *x-y* space. Second moments of the distribution, or rms sizes  $\sigma_{x,y}$  satisfy the *rms envelope equations* [10, 11]

$$\frac{d^2\sigma_x}{ds^2} + k_x(s)\sigma_x - \frac{\epsilon_{x,\text{rms}}^2}{\sigma_x^3} - \frac{K/2}{\sigma_x + \sigma_y} = 0 \quad (8)$$

$$\frac{d^2\sigma_y}{ds^2} + k_y(s)\sigma_y - \frac{\epsilon_{y,\text{rms}}^2}{\sigma_y^3} - \frac{K/2}{\sigma_x + \sigma_y} = 0 \quad (9)$$

where  $\sigma_x = \sqrt{x^2}$ ,  $\sigma_y = \sqrt{y^2}$  are the beam sizes, and  $\epsilon_{x,\text{rms}} = \sqrt{x^2 x'^2 - \bar{x}x'^2}$ ,  $\epsilon_{y,\text{rms}} = \sqrt{y^2 y'^2 - \bar{y}y'^2}$  are the rms emittances.

The fact that Eqs.(7,7) and (8,9) have the same form justifies the concept of equivalent uniform beams: rms envelopes of elliptically symmetric distributions are identical, independent of the details of the distributions. Therefore, a uniform distribution can be used to describe real beams. This concept does not hold exactly in the case of bunched beams, although it was shown that the dependence on the details of the distribution is weak [11]. The electric field of a 3-D ellipsoidal bunch with semi-axes  $r_x, r_y, r_z$  and uniform charge distribution, given by

$$\begin{aligned} E_x &= \frac{3I\lambda(1-f)}{4\pi\epsilon_0c(r_x+r_y)r_z} \frac{x}{r_x} \\ E_y &= \frac{3I\lambda(1-f)}{4\pi\epsilon_0c(r_x+r_y)r_z} \frac{y}{r_y} \\ E_z &= \frac{3I\lambda f}{4\pi\epsilon_0cr_xr_yr_z} \frac{z}{r_z} \end{aligned} \quad (10)$$

can be used to obtain the “equivalent” rms envelope equations for a bunched beam [1] (where  $r_x = \sqrt{5}\sigma_x$ , etc.),

$$\begin{aligned} \frac{d^2\sigma_x}{ds^2} + k_x(s)\sigma_x - \frac{\epsilon_{x,\text{rms}}^2}{\sigma_x^3} - \frac{K_3(1-f)}{(\sigma_x + \sigma_y)\sigma_z} &= 0 \\ \frac{d^2\sigma_y}{ds^2} + k_y(s)\sigma_y - \frac{\epsilon_{y,\text{rms}}^2}{\sigma_y^3} - \frac{K_3(1-f)}{(\sigma_x + \sigma_y)\sigma_z} &= 0 \\ \frac{d^2\sigma_z}{ds^2} + k_z(s)\sigma_z - \frac{\epsilon_{z,\text{rms}}^2}{\sigma_z^3} - \frac{K_3f}{\sigma_x\sigma_y} &= 0 \end{aligned} \quad (11)$$

where

$$K_3 = \frac{3qI\lambda}{20\sqrt{5}\pi\epsilon_0mc^3} \frac{1}{\beta^2\gamma^3}$$

is the 3-D space charge parameter,  $I = qNc/\lambda$  is the beam current averaged over the rf period,

$f$  is the geometric form-factor which is a function of the ellipsoid aspect-ratio parameter  $p = \gamma\sigma_z/\sqrt{\sigma_x\sigma_y}$ . For  $0.8 < p < 5$ , a useful approximation is  $f = 1/3p$  [12]. The Trace-3D code calculates beam envelopes for a bunched beam with linear space charge force [13].

The periodic solution of Eqs.(11) provides envelopes of a beam matched to the focusing channel. *Structure resonances* can occur when oscillation of the beam size deviation from the matched values is in resonance with the periodic focusing force for some beam envelope modes. To avoid the strongest resonance, the transverse zero-current phase advance per period should be kept below  $90^\circ$  [14].

The collective field of the beam exhibits a defocusing force on individual particles within the beam thus reducing the focusing action of external fields. The *tune depression*  $\eta$  quantifies the strength of the space-charge defocusing force. The longitudinal and transverse tune depressions are

$$\eta_z = \frac{\nu_z}{\nu_{z0}} = \sqrt{1 - \frac{fK_3}{\sigma_x\sigma_y\sigma_z} \left( \frac{L}{\nu_{z0}} \right)^2}$$

$$\eta_{x,y} = \frac{\nu_{x,y}}{\nu_{x0,y0}} = \sqrt{1 - \frac{(1-f)K_3}{(\sigma_x + \sigma_y)\sigma_x\sigma_y\sigma_z} \left( \frac{L}{\nu_{x0,y0}} \right)^2}$$

where  $L$  is the length of the focusing period,  $\nu_z, \nu_{x,y}$  are the phase advances per focusing period with space-charge, and  $\nu_{z0}, \nu_{x0,y0}$  are the zero-current phase advances.

**Halo and emittance growth** The nonlinear time-dependent space-charge forces in a beam with non-stationary distribution can cause the rms emittance to increase. The free-energy concept states that a stationary beam is in its lowest energy state and any deviation from equilibrium, or mismatch, provides a source of free energy which can transfer to an emittance increase. This concept provides an estimate of the upper limit for the emittance growth in a continuous beam [14],

$$\frac{\epsilon_f}{\epsilon_i} = \sqrt{1 + \frac{2h}{\eta^2}}$$

where  $\epsilon_i, \epsilon_f$  are the initial and final emittances (average for  $x$  and  $y$  for non-round beams),  $\eta$  is the tune depression, and  $h$  is a dimensionless free-energy parameter which can be calculated for each particular type of mismatch [14].

In the case of strong anisotropy between longitudinal and transverse planes in a bunched

beam, the degree of freedom with higher temperature can be a source of free energy for the emittance growth in the other plane. To prevent emittance redistribution the condition of *equipartition*,  $\epsilon_{nx}k_x = \epsilon_{ny}k_y = \epsilon_{nz}k_z$ , has to be met [14], where  $\epsilon_n$  is the normalized emittance, and  $k$  is the phase advance per unit length including space charge.

Another approach is to slow down the redistribution process by avoiding space charge coupling resonances. The so-called 'Hofmann diagram' is a useful tool for choosing an optimal working point trajectory when designing a high intensity linac [15].

Space-charge-induced emittance growth is often associated with the formation of low density tails, or halo, which can cause particle loss and is therefore especially undesirable for high intensity linacs. The principal halo-formation mechanism is a resonant interaction between the particles in the beam and a beam core that is oscillating because of mismatch. The particle-core model uses the envelope equation for calculating space-charge fields in a mismatched core and single particle equations for calculating individual particle motion in that field. In the case of a round continuous beam [16],

$$\frac{d^2r}{ds^2} + r - \frac{\eta^2}{r^3} - \frac{1-\eta^2}{r} = 0 \quad (12)$$

$$\frac{d^2x}{ds^2} + x - (1-\eta^2) \times \begin{cases} x/r^2 \\ 1/x \end{cases} = 0 \quad \begin{cases} |x| < r \\ |x| \geq r \end{cases} \quad (13)$$

where  $r$  is the ratio of the core radius to the matched core radius,  $x$  is the ratio of the particle displacement to the matched core radius, and  $s = k_0 z$  is the phase advance.

The maximum resonant particle amplitudes can be found by solving Eqs.(12, 13) numerically as a function of mismatch parameter  $\mu$ , defined as the ratio of the initial beam radius to the radius of the matched beam. In the range of tune depression  $0.5 < \eta < 0.9$  and mismatch parameter  $0.5 < \mu < 0.95$  and  $1.05 < \mu < 2$  the maximum amplitude is given by an approximated formula:  $x_{\max} \simeq 4(1 + |\ln \mu|)$  [16].

There is no complete theory to describe dynamics of non-stationary beam distributions analytically and numerical simulations should be used for detailed study. Several codes have been developed to calculate linac beam dynamics including space-charge forces using particle-in-cell methods. PARMILA [17], IMPACT [18],

TraceWin [19], and Track [20] are examples of more popular codes [21].

The principal challenge in modern high current proton linac design is the minimization of beam loss due to halo growth [22, 23, 24, 25].

## References

- [1] T. Wangler, RF Linear Accelerators, Wiley (2008)
- [2] P. Lapostolle, A. Septier, eds., Linear Accelerators, North-Holland (1970)
- [3] B. Schnizer, PA 2 (1971) 141
- [4] N. Pichoff, Proc. CERN Acc. School (2000)
- [5] L.W. Smith, R.L. Gluckstern, RSI 26 (1955) 220
- [6] M. Prome, in [2]
- [7] A. Carne et al, in [2]
- [8] P. Lapostolle, Los Alamos LA-11601-MS
- [9] I.M. Kapchinsky, V.V. Vladimirskey, Proc. Int. Conf. on High Energy Acc. (1959) p.274
- [10] P. Lapostolle, IEEE Trans. Nucl. Sci. NS-18 (1971) 1101
- [11] F.J. Sacherer, IEEE Trans. Nucl. Sci. NS-18 (1971) 1105
- [12] P.M. Lapostolle, CERN report AR/Int. SG/65-15 (1965)
- [13] K.R. Crandall, D.P. Rusthoi, Los Alamos LA-UR-97-886
- [14] M. Reiser, Theory and Design of Charged Particle Beams, Wiley (1994)
- [15] I. Hofmann, PRST-AB 6 (2003) 024202
- [16] T.P. Wangler et al, PRST-AB 1 (1998) 084201
- [17] H. Takeda, LANL LA-UR-98-4478
- [18] J. Qiang et al, J. Comp. Phys. 163 (2000) 434
- [19] D. Uriot et al, PAC 03, p. 3491
- [20] V. Aseev et al, PAC 05, p. 2053
- [21] R. Ryne, ICFA HB 08
- [22] Proc. 29th ICFA Adv. Beam Dyn. Wkshp, HALO03, AIP Conf Proc. 693
- [23] S. Nath et al, PAC 03, p.1515
- [24] F. Gerigk et al, LINAC 02, p.569
- [25] R. Duperrier, IPAC 10, p. 1246

### 2.4.7 Vlasov and Fokker-Planck Equations

B. Zotter, CERN

**Liouville theorem** The Vlasov Equation describes the evolution of a continuous distribution function  $\psi$  in phase space. It is essentially a collisionless Boltzmann equation, or an expression of the Liouville theorem, which states that, in the absence of collisions and dissipation, the local density in phase space must remain constant  $d\psi/dt = 0$ . Expressed in conjugate variables  $\vec{p}$

and  $\vec{q}$  [1],

$$\frac{\partial \psi}{\partial t} + \dot{\vec{q}} \cdot \frac{\partial \psi}{\partial \vec{q}} + \dot{\vec{p}} \cdot \frac{\partial \psi}{\partial \vec{p}} = 0 \quad (1)$$

Written in spatial and velocity coordinates  $\vec{x}, \vec{v}$ ,

$$\frac{\partial \psi}{\partial t} + \vec{v} \cdot \frac{\partial \psi}{\partial \vec{x}} + \frac{1}{m} \vec{F}(\vec{x}, t) \cdot \frac{\partial \psi}{\partial \vec{v}} = 0 \quad (2)$$

with  $d\vec{p}/dt$  replaced by the force on a particle of charge  $e$  and mass  $m$ . The Lorentz force  $\vec{F} = e(\vec{E} + \vec{v} \times \vec{B})$  should be obtained from the self-consistent EM fields.

**Linearized Vlasov equation** Consider 2-D phase space  $(p, q)$ , and write  $\psi$  as the sum of a time independent (stationary) part  $\psi_0(p, q)$  and a small time dependent (perturbation) part  $\psi_1(p, q, t)$ . For a Hamiltonian system,  $\dot{q} = \partial H/\partial p$  and  $\dot{p} = -\partial H/\partial q$ ,  $\psi_0$  satisfies

$$\frac{\partial \psi_0}{\partial q} \frac{\partial H}{\partial p} - \frac{\partial \psi_0}{\partial p} \frac{\partial H}{\partial q} = [\psi_0, H] = 0 \quad (3)$$

where the square brackets are Poisson brackets. It follows that  $\psi_0 = \psi_0(H)$ , and does not depend explicitly on  $p$  or  $q$ .

The linearization consists in neglecting terms of second order in small quantities,

$$\frac{\partial \psi_1}{\partial t} + \dot{q} \frac{\partial \psi_1}{\partial q} + \frac{1}{m} F_0 \frac{\partial \psi_1}{\partial p} = -\frac{1}{m} F_1 \frac{\partial \psi_0}{\partial p} \quad (4)$$

where  $F_0$  is the stationary part of the force (due to the applied external EM fields or force generated by  $\psi_0$ , e.g. through wake fields);  $F_1$  is the small, time-dependent part of the force generated by  $\psi_1$  through wake fields. If the perturbation is oscillatory  $\sim \exp(-i\omega t)$ , one can replace  $(\partial/\partial t) \rightarrow -i\omega$ . The equation permits the determination of  $\omega$  (complex). The distribution (and thus the beam) becomes unstable if  $\omega$  has a positive imaginary part exceeding radiation damping.

**Fokker-Planck equation** [2] In the presence of dissipative force or random noise, the phase space density is no longer conserved. This leads to the FP equation

$$\frac{\partial \psi}{\partial t} + \dot{\vec{q}} \cdot \frac{\partial \psi}{\partial \vec{q}} + \dot{\vec{p}} \cdot \frac{\partial \psi}{\partial \vec{p}} = 2\alpha\psi + D \frac{\partial^2 \psi}{\partial \vec{p}^2} \quad (5)$$

$\alpha, D$  = the damping and diffusion coefficients.

In case damping and noise come from synchrotron radiation in electron storage rings, we have [3, 4]

|          | Horiz.   | Vert.  | Long.   |
|----------|--|--|---|
| $(q, p)$ | $(-\frac{c}{\omega_x}x', x)$   | $(y, \frac{c}{\omega_y}y')$  | $(z, -\frac{\eta c}{\omega_s}\delta)$   |
| $\alpha$ | $\frac{U_0}{2ET_{\text{rev}}}$   | $\frac{U_0}{2ET_{\text{rev}}}$   | $\frac{U_0}{ET_{\text{rev}}}$   |
| $D$      | $\frac{\rho^2 D_u}{E^2 \nu_x^4}$   | $\frac{c^2 D_u}{2\omega_y^2 E^2 \gamma^2}$   | $\frac{\eta^2 c^2 D_u}{\omega_s^2 E^2}$   |
| rms      | $\sigma_x = \frac{c\sigma_{x'}}{\omega_x}$<br>$= \frac{\sqrt{C_q \gamma^2} \rho}{\nu_x^2}$ | $\sigma_y = \frac{c\sigma_{y'}}{\omega_y}$<br>$= \frac{\sqrt{C_q} \rho}{\sqrt{2} \nu_y}$ | $\sigma_\delta = \frac{\omega_s \sigma_z}{\eta c}$<br>$= \sqrt{\frac{C_q \gamma^2}{2\rho}}$ |

where  $U_0 = (4\pi/3)r_e m_e c^2 \gamma^4 / \rho$ ,  $D_u = \langle N u^2 \rangle / 2 = (55/48\sqrt{3})r_e \hbar m_e c^4 \gamma^7 / \rho^3$  ( $u$  photon energy),  $C_q = (55/32\sqrt{3})\hbar/m_e c$ ,  $E = m_e c^2 \gamma$ ,  $\eta$  slip factor,  $\rho$  bending radius. The rms size is given by  $\sigma_q^2 = \sigma_p^2 = D/2\alpha$ . In the  $x$ -motion,  $x$  is identified with  $p$  rather than  $q$  because quantum excitation occurs in  $x$  rather than  $x'$ .

## References

- [1] A. Hofmann, CERN 95-06, Vol.I (1995) p.259
- [2] J. Jowett, AIP Proc. 153 (1985) p.864
- [3] M. Sands, SLAC-121 (1970)
- [4] A.W. Chao, Lecture Notes in Phys. 296, Springer-Verlag (1986) p.51

### 2.4.8 Potential Well Effect

B. Zotter, CERN

Symbols  $I_b$  beam average current,  $n_B$  number of bunches in beam,  $V_{\text{rf}}$  rf voltage,  $\nu_{s0} = \omega_{s0}/\omega_0$ ,  $\nu_s = \omega_s/\omega_0$  zero-current and perturbed synchrotron tunes,  $b$  vacuum chamber pipe radius,  $\tau$  time displacement of particle from bunch center,  $h$  harmonic number,  $\eta$  slip factor,  $\varphi_s$  synchronous phase,  $\ell_0$  zero-current bunch length,  $\ell$  and  $\Delta E$  bunch length and energy spread,  $\xi = I_b/\dot{V}_{\text{rf}} = \alpha_p e I_b / \nu_s^2 E = 2\pi I_b / h V_{\text{rf}} \cos \varphi_s$  scaling parameter (dimension  $\Omega^{-1}$ ). We have

$$\nu_{s0} = \sqrt{\frac{h\eta V_{\text{rf}} \cos \varphi_s}{2\pi b^2 E/e}}, \quad \ell = \frac{|\eta| c}{\beta \omega_s} \frac{\Delta E}{E} \quad (1)$$

### Equation of motion [1]

$$\frac{d^2 \tau}{dt^2} + \omega_{s0}^2 \tau = \frac{\omega_{s0}^2 \xi}{2\pi} \left( \frac{Z_{\text{hom}}}{\omega_0} + \tau Z_{\text{pot}} \right) \quad (2)$$

## Sec.2.4: COLLECTIVE EFFECTS

where the higher-order-mode impedance  $Z_{\text{hom}}$  and the potential-well impedance  $Z_{\text{pot}}$  are

$$\begin{aligned} Z_{\text{hom}} &= \sum_{p=-\infty}^{\infty} \text{Re}Z(p\omega_0) J_0(p\omega_0 \hat{\tau}) \tilde{\lambda}(p\omega_0) \\ &\approx \sum_{p=-\infty}^{\infty} \text{Re}Z(p\omega_0) \tilde{\lambda}(p\omega_0) \\ Z_{\text{pot}} &= \sum_{p=-\infty}^{\infty} \text{Im}Z(p\omega_0) \frac{J_1(p\omega_0 \hat{\tau})}{\omega_0 \hat{\tau}/2} \tilde{\lambda}(p\omega_0) \\ &\approx \sum_{p=-\infty}^{\infty} \text{Im}Z(p\omega_0) p \tilde{\lambda}(p\omega_0) \end{aligned} \quad (3)$$

where  $\hat{\tau}$  is amplitude of synchrotron oscillations, and  $\tilde{\lambda}$  is Fourier transform of bunch line density. The second expressions are valid for small amplitude synchrotron oscillations. The perturbed synchrotron frequency is

$$\nu_s^2 = \nu_{s0}^2 \left( 1 - \frac{\xi}{2\pi} Z_{\text{pot}} \right) \quad (4)$$

while the center of the oscillation is shifted by  $\Delta\varphi = h\xi Z_{\text{hom}}/(2\pi)$ .

**Proton bunch lengthening** Without synchrotron radiation, the longitudinal emittance ( $\propto \ell \Delta E$ ) is invariant. It follows from Eq.(1) that

$$\ell^2 \omega_s = \ell_0^2 \omega_{s0} \quad (5)$$

For  $\ell > b$ , the impedance is mainly inductive with  $Z_{||}/n = i\omega_0 L = \text{constant}$  ( $n = \omega/\omega_0$ ), and for bunch with parabolic line density, [2]

$$\nu_s^2 = \nu_{s0}^2 \left( 1 + \frac{3}{\pi^2 B^3} \frac{I_b Z_c}{h V_{\text{rf}} \cos \varphi_s} \right) \quad (6)$$

$$Z_c = \left| \frac{Z_{||}}{n} \right| - \frac{g Z_0}{2\beta^2}$$

where  $B = \ell_{\text{full}}/2\pi R$  the bunching factor. The second term in  $Z_c$  is from direct space charge,  $g = 1 + 2 \ln(b/a)$  (Sec.2.4.5.1).

Eliminating the synchrotron frequency yields a quartic equation for the bunch-lengthening factor  $x = \ell/\ell_0$ ,

$$\nu_s^4 + \frac{\xi}{2\pi} \frac{3}{\pi^2 B_0^3} Z_c x - 1 = 0 \quad (7)$$

where  $B_0 = \ell_0/(2\pi R)$ .

**Electron bunch lengthening** For electrons, the energy spread is constant for low currents. The energy distribution is Gaussian. For linear motions the line density is also Gaussian. Eq.(1) yields

$$\sigma \omega_s = \sigma_0 \omega_{s0} \quad (\sigma = \sigma_z/c) \quad (8)$$

This leads to a cubic equation for  $x = \sigma/\sigma_0$ ,

$$x^3 - x + \frac{\xi}{2\pi} Z_c \frac{\sqrt{2\pi}}{\omega_0^3 \sigma_0^3} = 0 \quad (9)$$

Note: (i) Eq.(9) does not apply if  $\ell \ll b$  where resonant impedances dominate. (ii) Coherent synchrotron radiation can also produce bunch lengthening [3], although usually weaker than effect of the impedance.

**Turbulent threshold** The energy spread of an electron bunch remains unchanged with increasing current up to the turbulent threshold, where also the slope of bunch length versus currents increases. Dynamics in this regime is not completely understood. The bunch length above threshold can be expressed in terms of the scaling parameter  $\xi$ . For an impedance described by a power law  $Z_{||}(\omega) = (Z_{||})_0 (\omega/\omega_0)^a$ , one gets [4]

$$\frac{\sigma_z}{R} \propto \left( \left| \frac{Z_{||}}{n} \right|_0 \xi \right)^{1/(2+a)} \quad (10)$$

For an inductive wall  $a = 1$ , and  $\sigma \sim \xi^{1/3}$ . However, turbulence is mainly due to the high frequency part of the impedance evaluated at  $\omega_{\text{crit}} = c/\sigma_z$ . Then [5],

$$\left( \frac{\sigma_z}{R} \right)^3 = \frac{e}{\sqrt{2\pi} n_B} \frac{\alpha_p I_0}{\nu_{s0}^2 E/e} \left[ \left| \frac{Z_{||}}{n} \right|_{\text{crit}} + \left( \frac{Z_{||}}{n} \right)_{\text{pot}} \right] \quad (11)$$

The turbulent threshold current is

$$I_{\text{th}} = \frac{\sqrt{2\pi} k_b}{e} \frac{\nu_{s0}^2 E/e}{\alpha_p |Z_{||}/n|_{\text{crit}}} \sqrt{1 + \frac{(Z/n)_{\text{pot}}}{|Z/n|_{\text{crit}}}} \quad (12)$$

The bunch length at threshold  $\sigma_{\text{th}}/\sigma_0 = \left[ 1 + (Z/n)_{\text{pot}}/|Z/n|_{\text{crit}} \right]^{1/2}$ . Above threshold,

$$\begin{aligned} \nu_s &\propto \left[ 1 + \frac{(Z/n)_{\text{pot}}}{|Z/n|_{\text{crit}}} \right]^{-1/2} \\ \frac{\sigma_E}{\sigma_{E0}} &= \left[ 1 + \frac{(Z/n)_{\text{pot}}}{|Z/n|_{\text{crit}}} \right]^{-\frac{1}{2}} \frac{\sigma}{\sigma_0} \end{aligned} \quad (13)$$

**Haissinski equation** For given wake function, the equilibrium line density satisfies [6]

$$\lambda(\tau) = K \exp \left[ -U_0(\tau) - \xi \int_{-\infty}^{\tau} dt S(\tau+t) \lambda(t) \right] \quad (14)$$

where  $U_0(\tau) = \tau^2/2\sigma_0^2$  for a linear rf voltage (parabolic potential),  $S(\tau) = \int_0^{\tau} W_{||}(t) dt$  is the step-function wake potential with  $W_{||}$  the wake

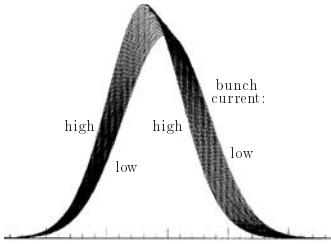


Figure 1: Electron bunch in resonator impedance for various currents.

function (Sec.3.2.1). The constant  $K$  is determined by normalization  $\int \lambda(t) dt = 1$ .

For a purely resistive impedance  $R_s$ , the Haissinski equation can be solved [7],

$$\lambda(\tau) = \frac{2\sqrt{(2/\pi)} \exp \tau^2/2\sigma_0^2}{\xi\sigma_0 [\operatorname{Coth}(\xi) - \operatorname{erf}(\tau/\sigma_0\sqrt{2})]} \quad (15)$$

where  $\xi \propto R_s I_b$ ; it can be obtained from the normalization  $\int d\tau \lambda(\tau) = 1$ . The solution corresponds to an asymmetric bell shape. For an inductive impedance  $S(\tau) = L\delta(\tau)$ , hence [8]

$$\lambda(\tau) = K \exp \left[ -\frac{\tau^2}{2\sigma_0^2} + \frac{\xi}{2} \lambda(\tau) \right] \quad (16)$$

Eq.(16) can be solved numerically and corresponds to bunch lengthening above transition. For a capacitive impedance, bunch shortens above transition.

For more realistic resonator impedances, solutions can be found by a perturbation algorithm [9]. See Fig.1. For other impedances, the Haissinski equation is integrated numerically. A divergence may occur for higher currents which has no physical significance [10]. The bunch deformation is important for determining instability thresholds [11].

## References

- [1] B. Zotter, CERN Report SPS/81-14 (1981)
- [2] S. Hansen et al, PAC 75, p.1381
- [3] C. Pellegrini, A. Sessler, Nuovo Cimento 53 B (1968) 198
- [4] A. Chao, J. Gareyte, PA 25 (1990) 229
- [5] A. Hofmann, J. Maidment, LEP-Note 168 (1979)
- [6] J. Haissinski, Nuovo Cimento 18 B (1973) 72
- [7] A. Ruggiero, PAC 77, p.1205
- [8] G. Besnier, B. Zotter, KEK Report 90-21 (1990) p.52
- [9] J. Hagel, B. Zotter, KEK Report 90-21 (1991) p.59
- [10] K. Bane, SLAC-PUB-5517 (1990)
- [11] K. Oide, KEK Report 90-21 (1990) p.64

### 2.4.9 Single-Bunch Instabilities in Circular Accelerators

B. Zotter, CERN

These instabilities can occur either in the longitudinal or in one of the transverse directions (usually more limiting vertically).

**Robinson instability** The detuning of a cavity by the induced current of a bunched beam leads to stability limits for the *tuning angle*  $\psi = \tan^{-1}[2Q_r(\omega_r - \omega_{rf})/\omega_r]$  [1, 2],

$$0 < \sin(2\psi) < \frac{V_{rf} \cos \varphi_s}{R_s I_0} \quad (1)$$

for machines operating below transition. The first condition requires tuning of the fundamental resonance  $\omega_r$  below the rf frequency  $\omega_{rf}$ . The second criterion limits the tuning angle as function of the rf voltage  $V_{rf}$ , the synchrotron phase angle  $\varphi_s$ , shunt impedance  $R_s$  and average current  $I_0$ . Up to now, the second limit has not been clearly observed. Above transition, the opposite inequalities hold. See also Sec.7.3.1.1.

Higher-order cavity modes may occur on either side of a harmonic  $p$  of the revolution frequency, contribute to damping, or drive instabilities with growth rates

$$\frac{1}{\tau_R} = \frac{2\omega_s I_0}{h V_{rf} \cos \varphi_s} \times \operatorname{Re} \{ Z_{||}[(p+\nu)\omega_0] - Z_{||}[(p-\nu)\omega_0] \} \quad (2)$$

The fundamental resonance is usually strongest, its correct tuning yields *Robinson damping*. The sum over  $p$  and all impedances determines beam stability.

**Negative mass instability** [3] Particles with  $v \approx c$  cannot travel much faster by acceleration but will increase their momentum and thus move with a larger radius of curvature. Above transition energy, particles which normally repel each other, seem to experience an attractive force. This is known as *negative mass effect* for its similarity with the formation of planetary rings with an otherwise attractive gravitational force outside the *Roche limit*. In uniform beams this may lead to self-bunching and particle loss when the current exceeds a threshold

$$I_{\text{thresh}} = F' \frac{E_0}{e} \frac{|\eta|}{\gamma Z/n} \left( \frac{\Delta p_{\text{FWHM}}}{m_0 c} \right)^2 \quad (3)$$

where  $\Delta p_{\text{FWHM}}$  is the momentum spread (full width at half maximum). The form factor  $F'$ , of order unity, depends on the particle distribution (see Tab.1).

**Resistive wall instability** Currents are induced in metallic vacuum chamber walls by the beam. Due to the finite resistivity of the walls, they extend behind the position of ultrarelativistic charges. Their EM fields act on charged particles arriving later and may increase their oscillation amplitudes, and cause instabilities in bunched or unbunched beams, in the longitudinal [4] or a transverse direction [5].

The skin depth in a metal of resistivity  $\rho_c$  and permeability  $\mu$  is  $\delta_s = \sqrt{2\rho_c/(\mu\omega)}$ . For metal walls thicker than  $\delta_s$ , the *resistive wall impedances* of a circular pipe of radius  $b$  are

$$\frac{Z_{||}}{n}(\omega) = \frac{\mu Z_0}{2\mu_0 b} \delta_s, \quad Z_{\perp}(\omega) = \frac{\mu R Z_0}{\mu_0 b^3} \delta_s \quad (4)$$

where  $2\pi R$  is accelerator circumference. Since  $Z_{||}/n \propto Z_{\perp} \propto \delta_s \propto \omega^{-1/2}$ , the largest impedance occurs at the lowest frequency. For transverse instabilities, this lowest frequency occurs at

$$\omega_{\min} = \omega_0 \min\{\nu - [\nu], [\nu + 1] - \nu\} \quad (5)$$

where  $[\nu]$  = integer part of the betatron tune  $\nu$ . See Eq.(11), Sec.2.4.10 for growth rate. This single spectral line has narrow band width, and thus oscillates for a long time and may cause coupled bunch instabilities. Impedances at all other betatron frequencies may be grouped into a wide frequency band. These fields decay rapidly, hence drive only single bunch instabilities such as head-tail modes discussed below.

Longitudinal coupled-bunch instabilities can also be caused by a resistive wall, but are usually weaker. In general, impedances due to all elements should be added to determine beam stability, as can be done with computer programs [6, 7].

**Microwave instability** In addition to potential well bunch lengthening (Sec.2.4.8), an increase of bunch current  $I_b$  can lead to longitudinal instabilities of a single bunch, independent of the presence of other bunches in the machine. [See also Eqs.(12-15), Sec.2.4.10] Because it is often accompanied by high-frequency signals, it is also called microwave instability. Oscillation frequencies change with amplitude due to non-linearities, hence this instability is usually self-limiting and only rarely leads to particle loss, but may reduce luminosity in colliders since bunches do not keep an optimum distribution.

Table 1: Form factors  $F'$  for various distributions.

| $F'$             | $\arg Z = \tan^{-1}(\text{Im } Z / \text{Re } Z)$ |       |          |                  |
|------------------|---|-------|----------|------------------|
| Distrib.         | $-\pi/2$  | 0     | $\pi/2$  | Remarks          |
| Parabolic        | 1.047   | 0.798 | 0        | sharp-edge       |
| 3/2 power        | 1.061   | 1.061 | 1.061    | Keil-Schnell     |
| Quartic          | 1.073   | 1.194 | 2.203    | rounded          |
| Gaussian         | 0.942   | 1.359 | $\infty$ | rounded          |
| Triangular       | 0   | 2.0   | 0        | sharp-edge       |
| Rounded triangle | 0.555   | 1.98  | 1.265    | with 1% rounding |

For long bunches, e.g. of proton beams, the impedance is mainly inductive and the threshold current is given by the *Boussard criterion* [8],

$$\frac{Z_{||}}{n} \leq F' \frac{E_0}{e} \frac{|\eta| \gamma}{I_b} \frac{\sigma_z}{c T_0} \left( \frac{\Delta p_{\text{FWHM}}}{p_{||}} \right)^2 \quad (6)$$

where  $E_0 = mc^2$ ,  $I_b = N_B e / T_0$ . This agrees with the (simplified) unbunched-beam or *Keil-Schnell criterion* [9],

$$\frac{Z_{||}}{n} \leq F' \frac{E_0}{e} \frac{|\eta|}{\gamma I_0} \left( \frac{\Delta p_{\text{FWHM}}}{m_0 c} \right)^2 \quad (7)$$

applied to the *local values of current and energy spread* in a bunch. The form factor  $F'$ , shown in Tab.1, depends on particle distribution and ratio of real to imaginary part of the impedance [10, 11].

For short bunches, the impedance seen by the beam is dominated by resonances at higher frequencies, and may lead to *turbulent bunch lengthening*. A tentative explanation for this is *longitudinal mode coupling* [12] (Sec.2.4.10).

Operating with  $\eta < 0$  can reduce bunch lengthening since the usually inductive impedance then shortens it [13]. However, theoretical and experimental investigations show that the turbulent threshold is lower with the shorter bunch, giving a larger energy spread.

**Head-tail instability** Short-range transverse wake fields, excited by particles at the head of a bunch, may excite oscillations at its tail. Synchrotron motion brings these particles again to the head and they continue to excite particles behind. These oscillations will grow (head-tail instability) if they add in phase due to a finite chromaticity  $\xi = \Delta\nu/\delta$  [14, 15] and if the growth rate exceeds radiation and Landau damping.

In the lowest mode,  $m = 0$ , all particles in a bunch oscillate in phase, which corresponds to a rigid dipole oscillation at the betatron frequency. The  $m = 0$  mode is unstable for  $\xi < 0$ . Since the natural chromaticity of circular machines is negative, one must correct it with sextupoles. *Bunch shape* modes with  $m \geq 1$  oscillate at synchrotron side-band frequencies  $\omega_\beta \pm m\omega_s$ . In the  $m = 1$  mode, particles at head and tail have opposite phases (when  $\xi = 0$ ), the bunch “toggles” about its middle, while higher modes ( $m \geq 2$ ) oscillate with  $m$  nodes over the bunch length. Modes with  $m \geq 1$  may become unstable for  $\xi > 0$ , but their growth rates are usually small and easily stabilized by damping except for large machines with very strong transverse wakes.

For short bunches, the growth rate of the  $m$ -th mode is (ignoring radial modes) [16]

$$\frac{1}{\tau_m} = \frac{N_B r_0 c \xi \omega_\beta \hat{z}}{2\pi^2 \gamma c \eta \nu_\beta^2} \times \int_0^\infty d\omega \operatorname{Re} Z_\perp(\omega) J_m\left(\frac{\omega \hat{z}}{c}\right) J'_m\left(\frac{\omega \hat{z}}{c}\right) \quad (8)$$

with  $N_B$  number of particles per bunch,  $\hat{z}$  the  $z$ -amplitude of synchrotron oscillation (airbag model),  $r_0$  classical particle radius,  $\omega_\beta$  the betatron frequency.

Example For resistive wall [15],

$$\frac{1}{\tau_m} = -\frac{\sqrt{2} J_m}{\pi^{5/2}} \frac{r_0 c}{\sqrt{\mu \sigma_c}} \frac{N_B \xi \sqrt{\hat{z}}}{\eta \gamma b^3 \nu_\beta} \quad (9)$$

with  $\mu$  permeability,  $\sigma_c$  conductivity,  $b$  chamber half-height. The factor  $J_m = \left( \int_0^{\pi/2} dx \sqrt{\sin x} \right) \left( \int_0^\pi d\psi \cos m\psi \sqrt{\sin \frac{\psi}{2}} \right)$  for  $m = 0, 1, 2$  is approximately 2.9, -0.57, -0.21. For negative chromaticity  $\xi < 0$ , the  $m = 0$  mode is strongly unstable, while the higher modes are weakly unstable for  $\xi > 0$ .

Example For the case with constant wake-function  $W_1(z) = -W_0$ ,

$$\frac{1}{\tau_m} = \frac{8}{\pi(4m^2 - 1)} \frac{N_B r_0 W_0 \xi \hat{z}}{c \eta \gamma T_0 Z_0 \nu_\beta} \quad (10)$$

**Longitudinal head-tail instability** This instability appears when the momentum compaction has an appreciable nonlinear dependence on energy. It has been seen at the CERN-SPS [17]. The growth rate of the instability can be written [18]

$$\frac{1}{\tau_m} = \frac{16m^2 \eta_1 N_B r_0}{3\eta_0 \gamma C Z_0} \int_{-\infty}^{\infty} d\omega \frac{\operatorname{Re} Z_\perp(\omega)}{\sigma^2} \quad (11)$$

$$\times [\sigma J_m(\sigma) J_{m+1}(\sigma) + (1-m) J_m^2(\sigma)]$$

where  $\sigma = \omega \hat{z} / (2c)$ ,  $\hat{z}$  is maximum  $z$ -amplitude of synchrotron oscillation of all particles (waferbag model),  $\eta = \eta_0 + \eta_1 \delta + \dots$  is the phase slip factor (Sec.2.3.11). For the case  $m = 1$  (dipole mode), Eq.(11) can be rewritten as

$$\frac{1}{\tau_1} = -\frac{c \eta_1 \hat{z}}{3\eta_0 C E} \frac{d\Delta E}{d\hat{z}} \quad (12)$$

where  $\Delta E$  is the energy loss of the beam bunch to the impedance. Its dependence on bunch length  $\hat{z}$  characterizes the growth rate of the longitudinal head-tail instability.

#### Transverse mode coupling instability (TMCI)

TMCI, also called *fast head-tail instability* or *strong head-tail instability*, occurs when the frequencies of two neighboring head-tail modes approach each other due to detuning with increasing current during accumulation. The original name *transverse turbulence* [19] is more appropriate for beams being injected into machine well above threshold, when a large number of modes may become coupled simultaneously.

Its threshold current is lowest in the  $y$ -plane since many elements are less high than wide,

$$I_{\text{thresh}} = \frac{2\pi \nu_s E / e}{\sum_i \beta_i Z_{\perp i}} F(\sigma_z) \quad (13)$$

The form factor  $F \approx 1$  for short bunches. For longer bunches, it increases proportionally to  $\sigma_z$ , and is essentially the ratio of the machine impedance to the effective impedance. For Gaussian bunches and broad-band resonator impedances,  $I_{\text{thresh}}$  can be expressed with the transverse loss or *kick factor*  $\kappa_\perp(\sigma_z)$  which eliminates the need for a bunch length correction factor,

$$I_{\text{thresh}} = \frac{C_1 f_s E / e}{\sum_i \beta_i \kappa_{\perp i}(\sigma_z)} \quad (14)$$

The constant  $C_1 \approx 8$ , but is often replaced by  $2\pi$  [11] as in Eq.(13). For more exact calculations of the threshold one should use computer programs solving for the coherent modes [20], including bunch lengthening [21] and potential-well deformations with current.

For the case of constant wake with  $W_1(z) = -W_0$  and a two-particle model beam bunch, the TMCI threshold occurs at

$$I_{\text{thresh}} = \frac{8}{\pi} \frac{\omega_\beta \omega_s (E/e)}{c W_0} \quad (15)$$

where  $\omega_{\beta,s}$  are the betatron and synchrotron frequencies.

## Sec.2.4: COLLECTIVE EFFECTS

TMCI is a severe limitation of single bunch current for large high-energy electron storage rings which have a high transverse impedance  $Z_{\perp} \approx (2R/b^2)Z/n$  because of their large radius  $R$  and the large number of rf cavities.

## References

- [1] K. Robinson, CEA Report CEAL-1010 (1964)
- [2] Yu. F. Orlov, E.K. Tarasov, JEPT 34(7), No.3, 339 (1958)
- [3] X. Nielsen, A. Sessler, Proc. Int. Conf. High Energy Acc. (1959) p.239
- [4] K. Neil, A. Sessler, RSI 36 (1965) 429
- [5] J. Laslett, K. Neil, A. Sessler, RSI 36 (1965) 436
- [6] A. Hofmann et al, PAC 79, p.1514
- [7] M. Zisman et al, LBL-21270 (1985)
- [8] D. Boussard, CERN/PS-BI (1972)
- [9] E. Keil, W. Schnell, CERN/ISR-TH/69-48 (1969)
- [10] A. Ruggiero, V. Vaccaro, CERN/ISR-RF/68-33 (1968)
- [11] B. Zotter, CERN Report ISR-GS/76-11 (1976)
- [12] F. Sacherer, PAC 77, p.1393
- [13] S.X. Fang et al, KEK 94-190 (1995)
- [14] C. Pellegrini, Il Nuovo Cimento 44 (1969) p.447
- [15] M. Sands, SLAC Report TN-69-8 and 10 (1969)
- [16] A.W. Chao, Physics of Collective Beam Instabilities in High Energy Accelerators, Wiley (1993)
- [17] D. Boussard, T. Linnecar, EPAC 90, p.1560
- [18] B. Chen, Thesis, U. Texas at Austin (1995)
- [19] R. Kohaupt, Int. Conf. Part. Acc. (1980) p.562
- [20] Y. Chin, CERN/SPS 84/2 (1984)
- [21] B. Zotter, CERN/LEP/AP 89-82 (1989)

### 2.4.10 Sacherer Formulae

*B. Zotter, CERN*

In longitudinal phase space  $x = r \cos \theta$ ,  $\dot{x} = \omega_{s0} r \sin \theta$ , a bunch shape mode  $\sim \exp(-im\theta)$  is specified by the azimuthal mode number  $m$  ( $m = 1$  dipole,  $m = 2$  quadrupole etc.), when only the lowest radial mode is considered. Its projection onto real space  $\lambda_m(x)$  is called line density mode. The coherent frequency  $\omega_m$  can be obtained from the (complex) coherent frequency shift  $\Delta\omega = \omega_{coh} - \omega_{inc}$ , if the incoherent oscillation frequency  $\omega_{inc}$  is known [1]. For  $n_B > 1$  bunches, the coupled bunch mode number  $0 \leq n < n_B$  gives the phase-shift  $2\pi n/n_B$  between oscillations of adjacent bunches.

**Longitudinal bunch oscillations** For  $n_B$  equally spaced and populated bunches, with

total beam current  $I_{tot} = n_B I_B$ , the coherent frequency shift of for  $m \geq 1$  is [2]

$$\Delta\omega_m = -i \frac{m \omega_s}{m+1} \frac{I_{tot} Z_{eff}^{(m)}}{3B^3 h V_T \cos \varphi_s} \quad (1)$$

The incoherent synchrotron frequency  $\omega_s = \nu_s \omega_0$  is a function of beam current,  $\omega_{s0}$  is its value for vanishing current,  $V_T = V_{rf} \omega_s^2 / \omega_{s0}^2$  = the total voltage seen by the beam,  $V_{rf}$  = applied rf voltage with harmonic number  $h = \omega_{rf}/\omega_0$ , and  $\varphi_s$  = stable phase angle. Above transition,  $90^\circ < \varphi_s < 180^\circ$  is needed for stability, hence  $\cos \varphi_s < 0$ . An inductance, for which  $Z_{eff} = -i\omega_0 L$ , then gives a positive coherent frequency shift, compensating almost exactly its negative incoherent frequency shift. The bunching factor  $B = n_B f_0 \tau_L$  is the ratio of bunch length  $L = \beta c \tau_L$  to bunch spacing  $2\pi R/n_B$ .

Assuming the modes do not depend on impedance or beam current, the effective impedance (Sec.3.2.5) can be expressed

$$Z_{eff}^{(m)} = \omega_0 \sum_{p=-\infty}^{\infty} \frac{Z_{||}(\omega_{mp})}{\omega_{mp}} H_m(\omega_{mp}) \quad (2)$$

where the spectral power density  $H_m(\omega) \propto |\tilde{\lambda}_m(\omega)|^2$  is normalized by  $\sum_p H_m(\omega_{mp}) = 1$ . The oscillation spectrum is  $\omega_{mp} = \omega_0(n_B p + n + m\nu_s)$ . The summation over all frequencies can be replaced by one over only positive harmonics with upper and lower sidebands, as illustrated in Fig.1 for the case of 4 Gaussian bunches in a resonator impedance.

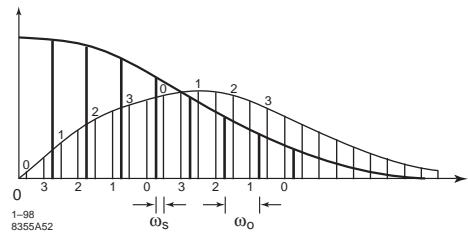


Figure 1: Effective impedance as summation over positive harmonics with upper and lower sidebands, weighted by spectral power density of a Gaussian. Indices 0,1,2,3 are multibunch mode number for  $n_B = 4$ .

**Proton bunches** have mostly a parabolic distribution, and **Legendre modes**  $\lambda_m \sim P_m(\tau)$  are eigenmodes in an inductive impedance. However, **sinusoidal modes**  $\lambda(\tau) \sim \sin[\pi m \omega_0 \tau]$  or

$\cos[\pi(m+1)\omega_0\tau/2]$  (for  $m$  odd or even) give better agreement with observation. Their spectral power density is

$$H_m(\omega) = C_m(m+1)^2 \frac{1 - (-)^{m+1} \cos \omega \tau_L}{\left[ \left( \frac{\omega \tau_L}{\pi} \right)^2 - (m+1)^2 \right]^2} \quad (3)$$

with  $C_m = 2n_B \omega_0 \tau_L / \pi^3$  for normalization. Proton bunches are usually long compared to the vacuum chamber diameter; the impedance is mainly inductive,  $Z_{||}(\omega)/n = -i\omega_0 L$  is constant and is equal to  $Z_{\text{eff}}$ , hence the frequency shift is mostly real. When the bunches are short compared to the rf wavelength, the coherent shift of the dominant dipole mode ( $m = 1$ ) is equal and opposite to the incoherent one, and thus  $\omega_{1,\text{coh}}$  remains constant with current.

Electron bunches usually have a Gaussian distribution, and oscillate in *Hermitian modes*,

$$H_m(\omega_{mp}) = D_m(\omega_{mp}\sigma)^{2m} \exp(-\omega_{mp}^2\sigma^2) \quad (4)$$

$D_m = \Gamma(m+1/2)/n_B\omega_0$  for normalization.

Neglecting potential well deformation and turbulence, the frequency spread due to a sinusoidal rf voltage is

$$S = \omega_s \frac{\pi^2}{16} \frac{1 + \frac{2}{3} \sin^2 \varphi_s}{\cos^2 \varphi_s} B^2 \quad (5)$$

The *effective spread* for a parabolic bunch is approximately  $mS/2(m+1)$ , while it is  $\sqrt{mS}/4$  for Gaussian bunches.

When modes with different mode numbers do not couple, an instability may occur if the total frequency shift  $|\Delta\omega|$  is larger than the effective frequency spread; Landau damping is lost, and if the growth rate  $1/\tau = -\text{Im}\Delta\omega$  exceeds radiation damping.

**Resonator impedance** With resonant frequency  $\omega_R$ , shunt impedance  $R_s$  and quality factor  $Q$ , the coherent frequency shift is

$$\frac{\Delta\omega_s}{\omega_0} = \frac{0.159 R_s I_0}{B h V \cos \varphi_s} F_m(\omega_R \tau_L) D(\alpha_d) \quad (6)$$

The form factor  $F_m(x)$  is proportional to  $x^m$  for short bunches, while it is  $64m/\pi x^2$  for long ones. The factor  $D$  depends on the attenuation  $\alpha_d = \frac{\pi\omega_R}{n_B Q \omega_0}$  between adjacent bunches,

$$D(x) = -i\alpha_d \left[ \frac{1}{1 - e^{x_+}} - \frac{1}{1 - e^{x_-}} \right] \quad (7)$$

where  $x_{\pm} = -2\pi i(n \pm \omega_R/\omega_0)/n_B - \alpha_d$ . The maximum value of  $|D|$  is about unity for narrow band, and approaches zero for wide band resonators.

**Transverse bunch oscillations** The transverse mode frequencies are  $\omega_{mp} = \omega_0(n_B p + n - m_t \nu_\beta + m \nu_s)$ . The transverse mode number  $m_t$  is usually taken to be unity (dipole oscillations), while the longitudinal one ( $m \geq 0$ ) should include the rigid bunch oscillation  $m = 0$ , which usually has the largest frequency shift and is most likely to become unstable [6, 7].

The coherent frequency shift of a bunch with total current  $I_{\text{tot}} = n_B I_B$  and energy  $E$  is [3, 4]

$$\Delta\omega_m = -\frac{iec}{m+1} \frac{I_{\text{tot}} Z_{\perp \text{eff}}^{(m)}}{4\pi\nu_\beta E B} \quad (8)$$

with the effective transverse impedance

$$Z_{\perp \text{eff}}^{(m)} = \sum_{p=-\infty}^{\infty} Z_{\perp}(\omega_{mp}) H_m(\omega_{mp} - \omega_\xi) \quad (9)$$

$H_m$  is evaluated at the mode frequency  $\omega_{mp}$  shifted by the *chromatic frequency*  $\omega_\xi = \xi\omega_0/\eta\nu_\beta$  where  $\xi = d\nu_\beta/d$ .

(i) For a constant impedance  $Z_{\perp \text{eff}} \approx Z_{\perp} H_m(\omega_\xi)$ , one obtains the unbunched beam result divided by the bunching factor, and by  $(m+1)$  since higher modes are more difficult to drive.

(ii) For a narrow-band impedance, when only a single resonance at frequency  $\omega_R$  contributes to the sum for the effective impedance, the coherent frequency shift becomes

$$\Delta\omega_m = -\frac{iec}{m+1} \frac{I_{\text{tot}} Z_{\perp}(\omega_R)}{4\pi\nu_\beta E_0} F'(\chi - \omega_R \tau_L) \quad (10)$$

where  $\chi = \omega_\xi \tau_L$  is the chromatic phase-shift across the bunch, and  $F'(\omega \tau_L) = H_m(\omega)/B$ .

(iii) The resistive wall impedance has a single large peak at  $\omega_{\min}$  [Eq.(5), Sec.2.4.9]. The narrow-band (or long range) contribution is approximately

$$\Delta\omega_m = -\frac{iec}{m+1} \frac{Z_{\perp}(\omega_0) I_{\text{tot}}}{4\pi\nu_\beta E_0 \tau_L} \sqrt{\frac{\pi}{n_B B}} \quad (11)$$

The short range contribution can be approximated by replacing the infinite sum for the effective impedance by an integral over all frequencies, and is usually negligible.

**Longitudinal mode coupling** Different oscillation modes can strongly couple when detuning makes two mode frequencies approach each other. A single bunch will become unstable for currents above a threshold given essentially by the unbunched beam expression applied to the peak current. For longitudinal oscillations, *turbulent*

bunch lengthening will occur until the bunch is stable again. This is also called *microwave instability* (Sec.2.4.9). Its threshold can be expressed in terms of matrix elements which are functions of impedance and beam current,

$$M_{mk} = -i \frac{m \omega_s}{m+1} \frac{I_{\text{tot}} Z_{\text{eff}}^{(m,k)}}{3B^3 h V_T \cos \varphi_s} \quad (12)$$

with the generalized effective impedance

$$Z_{\text{eff}}^{(m,k)} = \sum_{p=-\infty}^{\infty} \frac{Z(\omega_{mp})}{\omega_{mp}/\omega_0} H_{mk}(\omega_{mp}) \quad (13)$$

where  $H_{m,k}(\omega) = \tilde{\lambda}_m(\omega) \tilde{\lambda}_k^*(\omega)$  is the *cross power density*. The coherent frequency  $\omega$  is determined by the condition

$$\begin{vmatrix} \omega - m\omega_s - M_{m,m} & -M_{m,m+1} \\ -M_{m,m+1} & \omega - (m+1)\omega_s \\ -M_{m+1,m+1} & \end{vmatrix} = 0 \quad (14)$$

The threshold current is determined by

$$|M_{m,m+1}| = \frac{1}{2} |M_{m+1,m+1} - M_{m,m} - \omega_s| \quad (15)$$

To couple different azimuthal modes requires strong impedances at frequencies above the beam-pipe cutoff, which occur rarely in practice. However, impedances at lower frequencies can couple radial modes with the same azimuthal mode number  $m$ , when the bunch shape is asymmetric, e.g. due to potential well deformation in resistive impedances [8, 9]. Contrary to its transverse counterpart (TMCI, Sec.2.4.9), coupling of radial modes occurs at lower beam currents than coupling of different azimuthal modes.

## References

- [1] F. Sacherer, CERN/SI-BR/72-5 (1972)
- [2] F. Sacherer, CERN/MPS-BR/73-3 (1973); PAC 73, p.825
- [3] F. Sacherer, Proc. 9-th Int. Conf. on Part. Acc. (1974) p.347
- [4] F. Sacherer, Erice School (1976), CERN 77-13 (1977) p.198; CERN/PS-BR/76-21 (1976)
- [5] F. Sacherer, CERN/PS-BR/77-5 and 6 (1977); PAC 77, p.1393
- [6] A. Hofmann, K. Hübner, B. Zotter, PAC 79, p.1514
- [7] B. Zotter, CERN/LEP/89-74 (1989)
- [8] K. Oide, K. Yokoya, KEK-Preprint 10 (1990)
- [9] M. D'yachkov, R. Baartman, EPAC 94, p.1075

### 2.4.11 Landau Damping

K.Y. Ng, Fermilab

Wake force excites a number of collective waves in a beam and displaces it from its equilibrium position. These waves of the beam center exchange energy among themselves, resulting in growth in amplitude for some and damping for some. The spread in oscillation frequency accelerates the damping and decelerates the growth. This process is often called *Landau damping* [1, 2, 3, 4].

#### Transverse oscillation of unbunched beam

Consider a coasting beam of energy  $E_0$  and betatron frequency  $\omega_\beta$  under the influence of a transverse wake  $W_1$  (Sec.3.2.1). The transverse displacement  $y(\theta, t)$  of a particle at azimuthal angle  $\theta$  around the accelerator ring is

$$\begin{aligned} & \left[ \left( \frac{\partial}{\partial t} + \omega_0 \frac{\partial}{\partial \theta} \right)^2 + \omega_\beta^2 \right] y(\theta, t) \\ &= -\frac{e I_0}{\gamma m \beta c T_0} \int_{-\infty}^0 dt' W_1(t' - t) \langle y(\theta, t') \rangle \end{aligned} \quad (1)$$

where  $T_0 = 2\pi/\omega_0$  is the revolution period,  $\beta c$  the particle velocity,  $I_0$  is the average beam current, and  $\langle \dots \rangle$  implies averaging over all particles according to the distribution  $\rho(\omega_\beta)$ . Solution can be obtained in the harmonic-frequency ( $n\cdot\Omega$ ) space via the transformation  $(1/4\pi^2) \int d\theta dt e^{-in\theta+i\Omega t}$ . In the upper part of the  $\Omega$ -plane ( $\text{Im } \Omega > \omega_g$ , with  $\omega_g$  being the fastest growth rate of  $y(\theta, t)$ ),

$$\tilde{y}_n(\Omega) = \frac{2i\bar{\omega}_\beta \kappa \langle \tilde{y}_n(\Omega) \rangle}{\omega_\beta^2 - \hat{\omega}^2} - \frac{i(\Omega - 2n\omega_0)y_{n0}}{2\pi(\omega_\beta^2 - \hat{\omega}^2)} \quad (2)$$

where the initial beam displacement is  $y(\theta, 0) = \sum_n y_{n0} e^{in\theta}$ ,  $\dot{y}(\theta, 0) = 0$ ,  $\hat{\omega} = \Omega - n\omega_0$ ,  $\bar{\omega}_\beta = \langle \omega_\beta \rangle$ ,  $\kappa = ec I_0 Z_1^\perp(\Omega)/(2\bar{\omega}_\beta E_0 T_0)$ , and  $Z_1^\perp$  the transverse impedance (Sec.3.2.1). Physically, with  $\kappa = \kappa_R + i\kappa_I$ ,  $\kappa_I$  is the betatron frequency shift due to impedance and  $\kappa_R$  is the growth rate. Analytic continuation into the whole  $\Omega$ -plane gives

$$\begin{aligned} \langle \tilde{y}_n(\Omega) \rangle &= -\frac{i y_{n0} (\hat{\omega} - n\omega_0) \int_C \frac{\rho(\omega_\beta)}{\omega_\beta^2 - \hat{\omega}^2} d\omega_\beta}{H(\Omega)} \\ H(\Omega) &= 1 - i 2 \bar{\omega}_\beta \kappa \int_C d\omega_\beta \frac{\rho(\omega_\beta)}{\omega_\beta^2 - \hat{\omega}^2} \end{aligned} \quad (3)$$

with integration path  $C$  going below the pole at  $\hat{\omega}$  and above the pole at  $-\hat{\omega}$ . Back to the  $\theta-t$  space,

$$\langle y(\theta, t) \rangle = \sum_n \int_W d\Omega e^{i(n\theta - \Omega t)} \langle \tilde{y}_n(\Omega) \rangle \quad (4)$$

with path  $W$  above all poles. Thus the center of the beam consists of many harmonic waves at frequencies determined by the zeroes of  $H(\Omega)$ .

As an example, consider the Lorentzian distribution  $\rho(\omega_\beta) = (S_{\frac{1}{2}}/\pi)/[(\omega_\beta - \bar{\omega}_\beta)^2 + S_{\frac{1}{2}}^2]$ , where  $S_{\frac{1}{2}}$  is the HWHM. Since  $S_{\frac{1}{2}}/\bar{\omega}_\beta \ll 1$  and  $|\kappa|/\bar{\omega}_\beta \ll 1$ , keeping the lowest order, the solution simplifies to

$$\langle y(\theta, t) \rangle = \sum_{n=-\infty}^{\infty} y_{n0} \frac{\bar{\omega}_\beta - n\omega_0}{2\bar{\omega}_\beta} \times \left[ e^{in\theta - i(\bar{\omega}_\beta + n\omega_0 + \kappa_R)t} e^{-(S_{\frac{1}{2}} + \kappa_R)t} + \text{c.c.} \right] \quad (5)$$

which are betatron waves corresponding to betatron sidebands of the revolution harmonics. Here  $n > 0$  corresponds to fast waves, which are stable because  $\kappa_R > 0$ . For the slow waves with  $n < 0$ ,  $\kappa_R < 0$  and there is stability only when  $S_{\frac{1}{2}} > |\kappa_R|$ . We see that the growth initiated by  $Z_1^\perp$  is counteracted by the spread in betatron frequency.

Note that even when  $\langle y(\theta, t) \rangle$  is damped to zero, the displacements of individual particles are not. In practice, any small initial displacement of the beam center will be damped immediately if  $S_{\frac{1}{2}} > |\kappa_R|$ , ensuring that  $Z_1^\perp$  will stop driving the individual displacements. In other words, Landau damping nips any instability growth in the bud.

$H(\Omega) = 0$  is called the dispersion relation because it gives frequency  $\Omega$  as a function of harmonic number  $n$ . Since  $\pm\hat{\omega}$  are far apart, the denominator can be linearized to give

$$H(\Omega) = 1 - ik \int_C \frac{\rho(\omega_\beta) d\omega_\beta}{\omega_\beta - \hat{\omega}} = 0 \quad (6)$$

To obtain the stability contour,  $\hat{\omega}$  is considered to be real, but with a positive infinitesimal imaginary part added. With  $u = (\hat{\omega} - \bar{\omega}_\beta)/S_{\frac{1}{2}}$ , the dispersion relation is normalized to the HWHM spread,

$$1 + \frac{(\Delta\omega)_0}{S_{\frac{1}{2}}} \left[ f(u) + ig(u) \right] = 0 \quad (7)$$

where  $(\Delta\omega)_0 = \omega - n\omega_0 - \bar{\omega}_\beta = -ik$ . In fact,

$$(\Delta\omega)_0 = -\frac{ie\beta I_0}{2E_0 T_0} \beta_\perp Z_1^\perp (\bar{\omega}_\beta + n\omega_0) \quad (8)$$

is the  $Z_1^\perp$ -driven complex frequency shift in the absence of Landau damping (subscript 0) and  $\beta_\perp$  is the betatron function at the impedance.  $f(u) + ig(u)$  is called the *transfer function* and can be measured (Sec.7.4.17):

$$f(u) = \wp \int \frac{\hat{\rho}(v) dv}{v - u}, \quad g(u) = \pi \hat{\rho}(u) \quad (9)$$

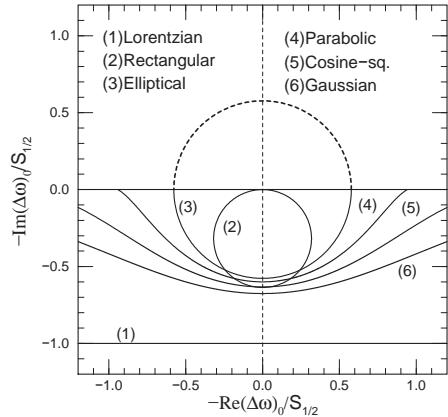


Figure 1: Plot of stability contours for various frequency distributions in the  $-(\Delta\omega)_0/S_{\frac{1}{2}}$ -plane.

where  $\wp$  denotes principal value,  $\hat{\rho}(v) = S_{\frac{1}{2}}\rho(\omega_\beta)$ ,  $v = (\omega_\beta - \bar{\omega}_\beta)/S_{\frac{1}{2}}$  so that  $v=1$  is at the HWHM.

The stability contour is the locus of  $-(\Delta\omega)_0/S_{\frac{1}{2}}$  as  $u$  is varied from  $-\infty$  to  $\infty$ . The beam is stable for points lying on the side of the locus containing the origin, or  $\text{Im } \Delta\omega_0 < S_{\frac{1}{2}}$  for the Lorentzian distribution. Let us consider next the generalized elliptical distribution  $\hat{\rho}(v) = (A_n/a_n)(1-v^2/a_n^2)^n$  when  $|v| \leq a_n$  and zero otherwise, where  $A_n = \Gamma(n+\frac{3}{2})/[\sqrt{\pi}\Gamma(n+1)]$  and the parameter  $a_n^2 = 1/(1-2^{-1/n})$  is so chosen that  $v=1$  is the HWHM. Note that  $n$  needs not be an integer. For example,  $n = 0, \frac{1}{2}, 1, \frac{3}{2}, \infty$  correspond, respectively, to the rectangular, elliptical, parabolic, tri-elliptical, and Gaussian distributions. In addition,  $n = 2.36$  reproduces closely the cosine-square distribution. Their stability contours are plotted in Fig.1.

Notice that all the contours, except the Lorentzian, intersect the  $-\text{Im}(\Delta\omega)_0/S_{\frac{1}{2}}$ -axis at roughly the same point  $-1/\sqrt{3}$ . A simplified stability criterion (Keil-Schnell criterion) can therefore be represented by a circle centered at the origin with radius equal to the intercept [6, 7],

$$|(\Delta\omega)_0| < \frac{1}{\sqrt{3}} S_{\frac{1}{2}} F_\perp \quad (10)$$

The form factor is  $F_\perp = \sqrt{3}$  for the Lorentzian distribution. For the generalized elliptical distribution,  $F_\perp = \sqrt{3}a_n/(\pi A_n)$ , or 1.103, 1, 1.040, 1.068, 1.097, and 1.174, respectively, for the rectangular, elliptical, parabolic, tri-elliptical, cosine square, and Gaussian distributions.

If the frequency spread  $S_{\frac{1}{2}}$  is due to a momentum spread  $\Delta\delta_{\frac{1}{2}}$ , the simplified stability criterion becomes

$$|Z_{\perp}(n\omega_0 + \bar{\omega}_{\beta})| < \frac{4\pi\beta E_0 \xi_{\text{eff}} \Delta\delta_{\frac{1}{2}}}{\sqrt{3}eI_0\beta_{\perp}} F_{\perp} \quad (11)$$

where  $\xi_{\text{eff}} = \xi - \eta(n + \nu_{\beta})$  with  $\xi$  the chromaticity,  $\nu_{\beta}$  the betatron tune, and  $\eta$  the slip factor. For a broad-band impedance rolling off at frequency  $\omega_c$ , the substitution  $n = \omega_c/\omega_0$  can be made.

**Transverse oscillation of a single bunch** When the bunch is very much shorter than the wavelength of the perturbing  $Z_1^{\perp}$ , the bunch can be viewed as a single macro-particle [3], oscillating transversely with frequency  $\Omega$ . The dispersion relation is the same as Eq.(6), but with  $\hat{\omega} = \Omega$ . The simplified stability criterion is still Eq.(10). The complex mode frequency shift due to wake effect in the absence of Landau damping is now

$$(\Delta\omega)_0 = (\Omega - \bar{\omega}_{\beta})_0 = \frac{e^2 N_B \beta_{\perp} \mathcal{W}}{2\beta^2 E_0 T_0} \quad (12)$$

$$\begin{aligned} \mathcal{W} &= \sum_{k=1}^{\infty} W_1(-kC) e^{i\bar{\omega}_{\beta} k T_0} \\ &= -\frac{i\beta}{T_0} \sum_{p=-\infty}^{\infty} Z_1^{\perp}(p\omega_0 + \bar{\omega}_{\beta}) \end{aligned} \quad (13)$$

with  $N_B$  the number of particles in the bunch.

When the length of the bunch is taken into account, there will be many more modes of transverse oscillation along the bunch. Head-tail instabilities can happen (Sec.2.4.9) and can be Landau damped when the growth rate  $\lesssim S_{\frac{1}{2}}$ .

Stronger impedance may cause two modes to merge, resulting in transverse mode-coupling instability (Sec.2.4.9). The growth is fast once above threshold and Landau damping usually does not help.

When the bunch is very much longer than the wavelength of  $Z_1^{\perp}$  and the growth rate is much faster than the synchrotron frequency, locally the bunch can be viewed as unbunched. Therefore the stability criterion of Eq.(10) applies [7]. However, we must replace  $I_0$  by the local peak current and  $\Delta\delta_{\frac{1}{2}}$  by the local momentum spread (Bousard criterion).

#### Longitudinal oscillation of unbunched beam

This case is unique in that there is no external focusing frequency. However, the longitudinal impedance  $Z_0^{\parallel}$  does alter the the particle's energy

and therefore its revolution frequency. The collective frequency shift,  $(\Delta\omega)_0$ , is similar to a synchrotron oscillation and is given, in the absence of Landau damping, by [5]

$$(\Delta\omega)_0^2 = (\Omega - n\omega_0)^2 = \frac{i\eta n I_0 \omega_0^2}{2\pi\beta^2 E_0} Z_0^{\parallel}(n\omega_0) \quad (14)$$

Landau damping arises from a spread in the revolution frequency. For a distribution  $\rho(\omega_0)$  with mean  $\bar{\omega}_0$ , the dispersion relation is [6]

$$1 - (\Delta\omega)_0^2 \int_C \frac{\rho(\omega_0) d\omega_0}{(\Omega - n\omega_0)^2} = 0 \quad (15)$$

where  $C$  goes below the pole at  $\Omega/n$ . The above, including the correct path of integration  $C$  can be derived by solving the Vlasov equation as an initial-value problem. At the onset of instability, integration by parts and normalization to the HWHM revolution frequency spread  $S_{\frac{1}{2}}$  result in

$$1 - \frac{(\Delta\omega)_0^2}{n^2 S_{\frac{1}{2}}^2} \int \frac{\hat{\rho}'(v) dv}{v - u - i\epsilon} = 0 \quad (16)$$

where  $\epsilon \rightarrow 0^+$ ,  $v = (\omega_0 - \bar{\omega}_0)/S_{\frac{1}{2}}$ ,  $nu = (\Omega - n\bar{\omega}_0)/S_{\frac{1}{2}}$ , and  $\hat{\rho}(v) = S_{\frac{1}{2}}\rho(\omega_0)$  so that  $v = 1$  is the HWHM. If the spread in  $\omega_0$  comes from the spread  $\Delta E$ , then  $S_{\frac{1}{2}} = (|\eta|\bar{\omega}_0/\beta^2)(\Delta E/E_0)^{\frac{1}{2}}$ . The dispersion relation becomes, in the  $U-V$  notation,

$$1 - \frac{2i}{\pi} \text{sgn}(\eta)(U + iV) \int \frac{\hat{\rho}'(v) dv}{v - u - i\epsilon} = 0 \quad (17)$$

$$U + iV = \frac{eI_0\beta^2 Z_0^{\parallel}(n\bar{\omega}_0)/n}{4|\eta|E_0(\Delta E/E_0)^{\frac{1}{2}}}$$

The stability contour can be traced by varying  $u$  from  $-\infty$  to  $\infty$ , and is depicted in Fig.2 for various distributions when  $\eta < 0$ .

Except for the Lorentzian distribution, all contours intersect the  $V$ -axis at roughly  $V_{in} = -1$ . A simplified stability criterion will therefore be the approximation of the contours as circles passing through  $V_{in}$ , or [6]

$$\left| \frac{Z_0^{\parallel}}{n} \right| < \frac{4|\eta|E_0}{e\beta^2 I_0} \left[ \frac{\Delta E}{E_0} \right]^2 F_{\parallel} \quad (18)$$

For the Lorentzian distribution, the form factor is  $F_{\parallel} = \pi/2$ . For the generalized elliptical distribution  $F_{\parallel} = \pi a_n^2/(4n+2)$ , which amounts to 1.047, 1.047, 1.061, 1.080, 1.097, and 1.133, respectively, for the elliptical, parabolic, tri-elliptical, cosine square,  $(1 - v^2/a_4^2)^4$ , and Gaussian distributions. The simplified stability criterion for the Gaussian distribution can also be written in the close form as  $|Z_0^{\parallel}/n| < 2\pi|\eta|E_0\sigma_E^2/(e\beta^2 I_0)$ , where  $\sigma_E$  is the fractional rms energy spread.

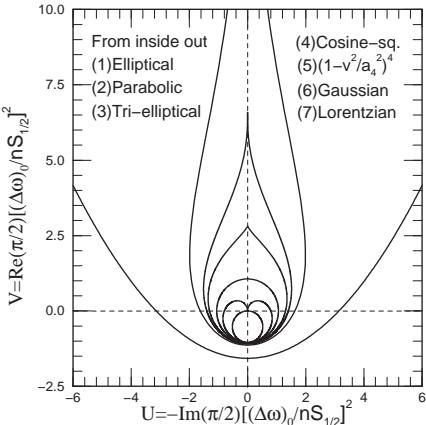


Figure 2: Stability contours for longitudinal unbunched beam in the  $\pi(\Delta\omega)_0^2/(2n^2S_{1/2}^2)$ -plane.

### Longitudinal oscillation of a single bunch

When the bunch is very much shorter than the wavelength of  $Z_0^\parallel$ , it can be approximated by a macro-particle [3]. In the presence of the synchrotron frequency  $\omega_s$ , the problem is similar to that of the transverse. As a result, the dispersion relation is still Eq.(6) but with  $\hat{\omega} = \Omega$ .  $(\Delta\omega)_0$  is the shift from the mean synchrotron frequency  $\bar{\omega}_s$  in the absence of Landau damping (potential-well distortion included, Sec.2.4.8),

$$(\Delta\omega)_0 = (\Omega - \bar{\omega}_s)_0 = \frac{e^2 N_B \eta c \mathcal{W}}{2\beta E_0 T_0 \bar{\omega}_s} \quad (19)$$

$$\begin{aligned} \mathcal{W} &= \sum_{k=1}^{\infty} e^{i\bar{\omega}_s k T_0} W_0''(-kC) \\ &= \frac{i}{C} \sum_{p=-\infty}^{\infty} (p\omega_0 + \bar{\omega}_s) Z_0^\parallel(p\omega_0 + \bar{\omega}_s) \end{aligned} \quad (20)$$

$W'_0$  being the longitudinal wake function and  $W''_0$  its derivative. Stability contours is given by Fig.1, and Eq.(10) remains a simplified stability criterion with  $S_{1/2}$  being the HWHM spread in  $\omega_s$ .

For finite bunch length, there will be many longitudinal modes of oscillation. Longitudinal head-tail instability can occur (Sec.2.4.9) but can be Landau damped by the spread in  $\omega_s$ .

When  $Z_0^\parallel$  is strong enough, two modes of oscillation can merge and longitudinal mode-coupling instability occurs (Sec.2.4.9). Once above threshold, the growth is fast and Landau damping usually does not help.

When the bunch is much longer than the wavelength of  $Z_0^\parallel$  and the synchrotron oscillation period is much longer than the instability growth time, the beam can be viewed as unbunched locally. Thus dispersion relation (17) and stability condition (18) apply [7]. However, the average current  $I_0$  must be replaced by the local peak current and the energy spread  $\Delta E$  by the local energy spread (Boussard criterion).

### References

- [1] L.D. Landau, J. Phys. USSR 10 (1946) 25
- [2] A. Hofmann, CERN 89-01 (1989)
- [3] A.W. Chao, Physics of Collective Beam Instabilities in High Energy Accelerators, Wiley (1993)
- [4] K.Y. Ng, Physics of Intensity Dependent Beam Instabilities, World Scientific (2006)
- [5] K.Y. Ng, AIP Conf. Proc. 184, Vol.2, p.1969; Fermilab Report FN-0438 (1986)
- [6] E. Keil, W. Schnell, CERN/TH-RF/69-48 (1969); V.K. Neil, A.M. Sessler, RSI 36 (1965) 429
- [7] D. Boussard, CERN Lab II/RF/Int 75-2 (1975)

### 2.4.12 Intrabeam Scattering and Touschek Effect

V. Lebedev, Fermilab

Coulomb scattering of charged particles in a beam results in an exchange of energy between different degrees of freedom. The total cross section of the two-particle scattering in vacuum diverges; however it has a finite value for collisions in plasma (or beam) due to field screening by other particles [1]. Usually two scattering regimes are considered: the single scattering when a rare single collision makes large change of particle momentum (Touschek effect), and the multiple scattering when multiple frequent collisions cause a diffusion. The former is usually responsible for creation of distribution tails and beam loss in electron machines, while the latter for changes in the distribution core. Although such approach is useful in many applications there are cases when it fails to deliver an accurate result.

### Multiple scattering in single component plasma

The temperature exchange in plasma is driven by interplay of friction and diffusion [2]. In uniform nonrelativistic one-component plasma with Gaussian distribution for each degree of freedom but different rms velocities

$$\left( \sigma_{vk} \equiv \sqrt{v_k^2}, k = (x, y, z) \right) \text{ it results in [3],}$$

$$\frac{d}{dt} \begin{pmatrix} \sigma_{vx}^2 \\ \sigma_{vy}^2 \\ \sigma_{vz}^2 \end{pmatrix} = \frac{(2\pi)^{3/2} nr_0^2 c^4 L_c}{\sqrt{\sigma_{vx}^2 + \sigma_{vy}^2 + \sigma_{vz}^2}} \times \begin{pmatrix} \Psi(\sigma_{vx}, \sigma_{vy}, \sigma_{vz}) \\ \Psi(\sigma_{vy}, \sigma_{vz}, \sigma_{vx}) \\ \Psi(\sigma_{vz}, \sigma_{vx}, \sigma_{vy}) \end{pmatrix} \quad (1)$$

where  $r_0$  is the particle classical radius,  $n$  is the plasma density,  $L_c = \ln(\rho_{\max}/\rho_{\min})$  is the Coulomb logarithm,

$$\rho_{\min} = r_0 c^2 / \bar{v^2}, \quad \rho_{\max} = \sqrt{\bar{v^2} / 4\pi n r_0 c^2} \quad (2)$$

where  $\bar{v^2} = \sigma_{vx}^2 + \sigma_{vy}^2 + \sigma_{vz}^2$ . Function  $\Psi(x, y, z)$  can be expressed through the symmetric elliptic integral of the second kind,  $R_D(x, y, z)$ ,

$$\Psi(x, y, z) = \frac{\sqrt{2}r}{3\pi} [y^2 R_D(z^2, x^2, y^2) + z^2 R_D(x^2, y^2, z^2) - 2x^2 R_D(y^2, z^2, x^2)] \quad (3)$$

$$R_D(x, y, z) = \frac{3}{2} \int_0^\infty \frac{dt}{\sqrt{(t+x)(t+y)(t+z)^3}}$$

where  $r = \sqrt{x^2 + y^2 + z^2}$ ;  $x, y, z \geq 0$ . An algorithm for fast numerical evaluation of  $R_D(x, y, z)$  is discussed in [4].  $\Psi(x, y, z)$  is chosen so that it depends on the ratios of its variables but not on  $r$ . It is symmetric relative to the variables  $y$  and  $z$ , and is normalized so that  $\Psi(0, 1, 1) = 1$ . The energy conservation yields  $\Psi(1, 0, 1) = \Psi(1, 1, 0) = -1/2$  and  $\Psi(x, y, z) + \Psi(y, z, x) + \Psi(z, x, y) = 0$ . In thermal equilibrium  $\Psi(1, 1, 1) = 0$ . If the first parameter is zero it can be approximated to  $\sim 0.5\%$  accuracy by

$$\Psi(0, x, y) \approx 1 + \frac{\sqrt{2}}{\pi} \ln \left( \frac{x^2 + y^2}{2xy} \right) - 0.055 \left( \frac{x^2 - y^2}{x^2 + y^2} \right)^2$$

The asymptotics are

$$\Psi(x, y, z) = \begin{cases} \frac{1}{\sqrt{2}\pi} \left( \ln \left( \frac{8r^2}{x^2 + y^2} \right) - \frac{3z}{r} \right), & x, y \ll z \\ \frac{\sqrt{2}}{\pi} \left( \ln \left( \frac{y^2 + z^2}{8r^2} \right) + \frac{3x}{r} \right), & y, z \ll x \end{cases} \quad (4)$$

Eq.(1) is obtained in the logarithmic approximation and is justified if  $\ln(\rho_{\max}/\rho_{\min}) \gg 1$  where the plasma perturbation theory can be used [5]. It describes the temperature relaxation in a wide class of electron and ion beams. Particularly, in the case of electrostatic acceleration the beam

longitudinal temperature is much smaller than the transverse one and its growth rate in the course of beam transport is [6] ( $\sigma_{vz} \ll \sqrt{\bar{v^2}}$ )

$$\frac{dT_z}{dt} \equiv m \frac{d\sigma_{vz}^2}{dt} = \frac{(2\pi)^{3/2} nmr_0^2 c^4 L_c}{\sqrt{\bar{v^2}}} \quad (5)$$

The temperature relaxation is suppressed for a beam moving along strong solenoidal magnetic field [7].

Note that Eq.(1) is not self-consistent; i.e. it implies that the initial distribution remains Gaussian; but an interplay of the diffusion and friction due to multiple scattering does not allow for the beam distribution to remain Gaussian during the process. Additionally the single scattering creates non-Gaussian tails. However deviations from the Gaussian distribution are comparatively small and Eq.(1) represents a good approximation in most practical cases.

**Multiple scattering in a storage ring** Although the beam focusing and finite dispersion in circular accelerators significantly complicate equations they do not change the essence of the process. However the time dependent nature of transverse focusing results in an absence of thermal equilibrium and a continuous growth of 3-D emittance. In the absence of  $x$ - $y$  coupling for a bunched beam nonrelativistic in the beam frame (BF) the emittance growth rates can be expressed in a compact matrix form

$$\frac{d\varepsilon_k}{dt} = \frac{Nr_0^2 c^2 / \sqrt{\varepsilon_x \varepsilon_y}}{4\sqrt{2}\sigma_z \beta^2 \gamma^4} \left\langle \frac{L_c \sum_{i,j=1}^3 \mathbf{B}_{ij}^k \mathbf{R}_{ij}}{\sqrt{\beta_x \beta_y F_D \text{tr}(\Sigma)}} \right\rangle_s \quad (6)$$

coinciding with the results of [4, 8, 9] for the case of zero vertical dispersion. Note that the derivatives of Courant-Snyder parameters omitted in [10] can significantly change results if optics has large  $\beta$ -function variations. Here  $\langle \dots \rangle_s$  denotes averaging over the machine circumference,  $\beta$  and  $\gamma$  are the relativistic factors;  $N$  is the number of particles per bunch;  $\Sigma \equiv [\bar{v_i v_j}]$  is the matrix of the second moments of local velocity distribution in the BF,

$$\Sigma = (\gamma \beta c)^2 \mathbf{G}^t \boldsymbol{\Xi}^{-1} \mathbf{G} \quad (7)$$

$\boldsymbol{\Xi}$  is the bilinear form of particle angles  $(\theta_x, \theta_y, \Delta p/p)$ ,

$$\boldsymbol{\Xi} = \begin{pmatrix} \beta_x/\varepsilon_x & 0 & -\beta_x \Phi_x/\varepsilon_x \\ 0 & \beta_y/\varepsilon_y & -\beta_y \Phi_y/\varepsilon_y \\ -\beta_x \Phi_x/\varepsilon_x & -\beta_y \Phi_y/\varepsilon_y & \boldsymbol{\Xi}_{33} \end{pmatrix} \quad (8)$$

$$\Xi_{33} = 1/\sigma_p^2 + A_x/\varepsilon_x + A_y/\varepsilon_y$$

$$\Phi_x = D'_x + \frac{\alpha_x D_x}{\beta_x}, \quad \Phi_y = D'_y + \frac{\alpha_y D_y}{\beta_y}$$

$$A_x = \frac{D_x^2 + (\beta_x \Phi_x)^2}{\beta_x}, \quad A_y = \frac{D_y^2 + (\beta_y \Phi_y)^2}{\beta_y}$$

$$F_D = 1 + D_x^2 \sigma_p^2 / (\varepsilon_x \beta_x) + D_y^2 \sigma_p^2 / (\varepsilon_y \beta_y)$$

$\sigma_z$  is the rms bunch length;  $\beta_x$ ,  $\beta_y$ ,  $\alpha_x$ , and  $\alpha_y$  are Courant-Snyder parameters;  $D_x$ ,  $D_y$ ,  $D'_x$  and  $D'_y$  are dispersions and their derivatives;  $\varepsilon_x$ ,  $\varepsilon_y$  and  $\varepsilon_z = \sigma_z \sigma_p$  are the unnormalized transverse and longitudinal rms beam emittances;  $\sigma_p$  is the relative rms momentum spread; matrices  $\mathbf{B}^k$  and  $\mathbf{G}$  are

$$\mathbf{B}^x = \begin{pmatrix} \beta_x & 0 & -\Phi_x \beta_x \\ 0 & 0 & 0 \\ -\Phi_x \beta_x & 0 & A_x \end{pmatrix}$$

$$\mathbf{B}^y = \begin{pmatrix} 0 & 0 & 0 \\ 0 & \beta_y & -\Phi_y \beta_y \\ 0 & -\Phi_y \beta_y & A_y \end{pmatrix}$$

$$\mathbf{B}^z = \text{diag}(0, 0, \beta_z), \quad \mathbf{G} = \text{diag}(1, 1, 1/\gamma)$$

the function  $\text{diag}(\cdot)$  makes a diagonal matrix from a vector;  $\beta_z = \sigma_z/\sigma_p$  is the longitudinal  $\beta$ -function;

$$\mathbf{R} = (\mathbf{G}^{-1})^t \mathbf{T} \Psi_{IBS} (\mathbf{T}^t \Sigma \mathbf{T}) \mathbf{T}^t \mathbf{G}^{-1} \quad (9)$$

a rotational matrix  $\mathbf{T}$  reduces  $\Sigma$  to its diagonal form  $\sigma = \mathbf{T}^t \Sigma \mathbf{T}$ ; and the matrix function  $\Psi_{IBS}(\sigma)$  with its argument of a diagonal matrix  $\sigma$  is

$$\Psi_{IBS}(\sigma) = \text{diag}(\Psi(\sqrt{\sigma_{11}}, \sqrt{\sigma_{22}}, \sqrt{\sigma_{33}}), \Psi(\sqrt{\sigma_{22}}, \sqrt{\sigma_{33}}, \sqrt{\sigma_{11}}), \Psi(\sqrt{\sigma_{33}}, \sqrt{\sigma_{11}}, \sqrt{\sigma_{22}}))$$

The Coulomb logarithm is computed similarly to the plasma case with the following correction affecting the value of maximum impact parameter of Eq.(2),

$$\rho_{\max} = \min\left(\sigma_{\min}, \gamma \sigma_z, \sqrt{\text{tr}(\Sigma) / 4\pi n r_0 c^2}\right)$$

where

$$2\sigma_{\min}^2 = \sigma_x^2 + \sigma_y^2 - \sqrt{(\sigma_x^2 - \sigma_y^2)^2 + 4D_x^2 D_y^2 \sigma_p^4}$$

$$\sigma_x^2 = \varepsilon_x \beta_x + D_x^2 \sigma_p^2, \quad \sigma_y^2 = \varepsilon_y \beta_y + D_y^2 \sigma_p^2$$

The above equations can be used for a coasting beam with substitutions  $\sigma_z \rightarrow L/(2\sqrt{\pi})$  and  $\mathbf{B}_z = \text{diag}(0, 0, 2)$  implying that  $d\varepsilon_3/dt \rightarrow$

$d\sigma_p^2/dt$ . Note that the factor of 2 in  $B_z$  reflects the absence of the synchrotron motion, taken into account in Eq.(6).

In many practical applications the longitudinal temperature in the BF is much smaller than the transverse one ( $\theta_{\parallel}/\gamma \ll \theta_{\perp}$ ) and the vertical dispersion can be neglected. Eq.(6) (bunched beam) then can be reduced to

$$\frac{d}{dt} \begin{pmatrix} \varepsilon_x \\ \varepsilon_y \\ \sigma_p^2 \end{pmatrix} = \frac{Nr_0^2 c}{4\sqrt{2}\beta^3 \gamma^3 \sigma_z} \times \left\langle \frac{L_c}{\sigma_x \sigma_y \theta_{\perp}} \begin{pmatrix} \Psi_x(\theta_x, \theta_y) \\ \Psi_y(\theta_x, \theta_y) \\ \Psi(0, \theta_x, \theta_y) \end{pmatrix} \right\rangle_s \quad (10)$$

where

$$\theta_{\perp} = \sqrt{\theta_x^2 + \theta_y^2}, \quad \theta_x^2 = \frac{\varepsilon_x}{\beta_x} \left( 1 + \frac{\sigma_p^2 (\Phi_x \beta_x)^2}{\sigma_x^2} \right)$$

$$\theta_y^2 = \frac{\varepsilon_y}{\beta_y}, \quad \Psi_y(\theta_y, \theta_x) = \Psi(\theta_y, \theta_x, 0) \frac{\beta_y}{\gamma^2}$$

$$\Psi_x(\theta_x, \theta_y) = \Psi(0, \theta_x, \theta_y) A_x + \Psi(\theta_x, \theta_y, 0) \frac{\beta_x}{\gamma^2}$$

For an ultrarelativistic machine,  $\gamma \gg Q_x$ , the terms with  $\beta_{x,y}/\gamma^2$  are small and can be neglected.

Similar to the suppression of IBS in an electron beam by magnetic field the multiple IBS is suppressed if  $\sigma_p$  becomes so small that the collision time becomes comparable to the betatron frequency [11].

**Touschek effect** Single scatterings become important when there is a large difference between rms velocities of different degrees of freedom in the BF. In this case a single scattering can result in a momentum transfer significantly exceeding the rms value of the coldest degree of freedom thus creating non-Gaussian tails and particle loss. For the case when the longitudinal temperature is much smaller than the transverse ones it is called the Touschek effect. It was first investigated in electron storage rings for a flat beam and nonrelativistic energies in the BF [12, 13]. The calculations were extended to arbitrary energies [14], round beams [15] and the dispersion was included [16]. A general case for an arbitrary ratio of transverse emittances and arbitrary energies was considered in [17]. The lifetime  $T$  of the particles having Gaussian distributions for all degrees of

freedom is [17]

$$\begin{aligned} \frac{1}{T} &= \frac{Nr_0^2c}{8\sqrt{\pi}\beta^2\gamma^4\varepsilon_x\varepsilon_y\sigma_z\sigma_p} \times \\ &\left\langle \frac{1}{\sqrt{\Xi_{33}}} \int_{u_m}^{\infty} \left[ \left(2 + \frac{1}{u}\right)^2 \left(\frac{u/u_m}{1+u} - 1\right) \right. \right. \\ &+ 1 - \sqrt{\frac{1+u}{u/u_m}} - \frac{1}{2u} \left(4 + \frac{1}{u}\right) \ln \left(\frac{u/u_m}{1+u}\right) \left. \right] \\ &\left. e^{-B_1 u} I_0(B_2 u) \frac{\sqrt{u}du}{\sqrt{1+u}} \right\rangle_s \quad (11) \end{aligned}$$

where  $u_m = \beta^2(\Delta p_m/p)^2$ ,  $\Delta p_m$  is the maximum stable momentum deviation and we assume it does not depend on a betatron motion amplitude,

$$B_1 = \frac{1}{2\beta^2\gamma^2} \left[ \frac{\beta_x}{\varepsilon_x} \left(1 - \frac{\beta_x\Phi_x^2}{\Xi_{33}\varepsilon_x}\right) + \frac{\beta_y}{\varepsilon_y} \left(1 - \frac{\beta_y\Phi_y^2}{\Xi_{33}\varepsilon_y}\right) \right]$$

$$B_2 = \sqrt{B_1^2 - \frac{\beta_x\beta_y}{\beta^4\gamma^4\varepsilon_x\varepsilon_y\Xi_{33}} \left( \frac{1}{\sigma_p^2} + \frac{D_x^2}{\varepsilon_x\beta_x} + \frac{D_y^2}{\varepsilon_y\beta_y} \right)}$$

and  $I_0(x)$  is the modified Bessel function. For  $D_y = 0$  (planar orbit)  $B_{1,2}$  simplify to

$$B_{1,2} = \frac{1}{2\beta^2\gamma^2} \left| \frac{\beta_x}{\varepsilon_x \left(1 + (\sigma_p\beta_x\Phi_x/\sigma_x)^2\right)} \pm \frac{\beta_y}{\varepsilon_y} \right|$$

For flat beams  $\varepsilon_x/\beta_x \gg \varepsilon_y/\beta_y$  and non-relativistic energies in the BF ( $\gamma^2\varepsilon_x/\beta_x \ll 1$ ) the Bessel function can be expanded for large arguments and one obtains

$$\frac{1}{T} = \frac{Nr_0^2c}{8\pi\beta^3\gamma^5\sigma_z} \left\langle \frac{(\beta_x/\varepsilon_x)^{\frac{3}{2}} C(\varepsilon_m)}{\left[1 + \left(\frac{\sigma_p\beta_x\Phi_x}{\sigma_x}\right)^2\right]^{\frac{3}{2}}} \right\rangle_s \quad (12)$$

with

$$\begin{aligned} C(x) &= -\frac{3}{2}e^{-x} + \int_{\varepsilon}^{\infty} \left(1 + \frac{3x}{2} + \frac{x}{2} \ln \frac{u}{x}\right) e^{-u} \frac{du}{u} \\ \varepsilon_m &= \frac{\Delta p_m^2}{p^2} \frac{\beta_x}{\gamma^2\varepsilon_x \left[1 + (\sigma_p\beta_x\Phi_x/\sigma_x)^2\right]} \end{aligned}$$

**IBS in hadron colliders** The radiation damping in electron synchrotrons typically results in the rms momentum spread being much smaller than the rf bucket size,  $\sigma_p \ll \Delta p_m/p$ . It has been

assumed in the Touschek lifetime, as well as in the IBS growth rates of Eq.(6) implying a linear rf waveform, i.e. a short bunch. However it is usually not the case in hadron colliders where the bunch fills most of the longitudinal acceptance. Initially the beam is well inside of the rf bucket and the beam loss is dominated by single scatterings. Later, the multiple scatterings result in the beam distribution reaching the rf bucket boundary leading to domination of multiple scattering in the particle loss. For the pancake velocity distribution in the BF  $\sigma_p/\gamma \ll \theta_{\perp}$  (which is always the case for colliders) the evolution of longitudinal distribution and the particle loss from the rf bucket due to multiple scattering can be described by a diffusion equation

$$\frac{\partial f}{\partial t} = \frac{1}{2} \frac{\partial}{\partial I} \left( I \frac{D(I)}{\omega(I)} \frac{\partial f}{\partial I} \right) \quad (13)$$

where  $I$  and  $\omega(I)$  are the action and frequency of synchrotron motion. For a single harmonic rf the dimensionless Hamiltonian and the diffusion are

$$H = \frac{\hat{p}^2}{2} + 2 \left( \sin \frac{\varphi}{2} \right)^2 \quad (14)$$

$$D(I) = \frac{\tilde{A}L_c}{\pi I \omega(I)} \oint \hat{p}^2 n(\varphi) d\psi$$

$$\tilde{A} = 2\sqrt{2\pi} \frac{\pi^3 N r_0^2 \beta^2 f_{RF}^3}{\beta^6 \gamma^3 c^2} \left\langle \frac{\Psi(0, \theta_x, \theta_y)}{\sigma_x \sigma_y \sqrt{\theta_x^2 + \theta_y^2}} \right\rangle_s$$

where  $n(\varphi) = \int f(I(\varphi, \hat{p})) d\hat{p}$  is the linear density normalized so that at the process beginning  $2\pi \int f(I) dI = \int_{-\pi}^{\pi} n(\varphi) d\varphi = 1$ ,  $\psi$  is the motion phase ( $d\psi = \omega(I)dt$ ), and  $f_{RF}$  is the rf frequency. An integration of Eq.(13) for the linear rf and Gaussian distribution yields the momentum growth rate of Eq.(10). To take into account the single scattering one needs to replace Eq.(13) by integro-differential equation [18]

$$\frac{\partial f(I, t)}{\partial t} = \tilde{A} \int_0^{\infty} W(I, I') (f(I', t) - f(I, t)) dI' \quad (15)$$

The kernel is a symmetric function,  $W(I, I') = W(I', I)$ , which for  $I' \geq I$  can be expressed as

$$\begin{aligned} W(I, I') &= \frac{\omega\omega'}{\pi (H(I) - H(I'))^3} \\ &\times \left[ (H(I) - H(I')) \int_0^{a(I)} n(\varphi) \frac{d\varphi}{\hat{p}} + 2 \int_0^{a(I)} n(\varphi) \hat{p} d\varphi \right] \end{aligned}$$

where  $a(I)$  is the motion amplitude. This equation has to be solved numerically together with the equations describing the evolution of transverse emittances.

**Intrabeam stripping** There is another mechanism of particle loss related to the intrabeam collisions of  $H^-$  ions. It results in stripping one of two  $H^-$  ions in a binary collision. It was first observed in the CERN antiproton ring LEAR operating with  $H^-$  ions during its commissioning [19]. It was recently found that the intrabeam stripping is a major mechanism of particle loss in the superconducting part of the SNS linac making almost uniform relative loss of  $\sim 10^{-6} \text{ m}^{-1}$  [20]. The particle loss cross section has a flat top of  $\sim 4 \times 10^{-15} \text{ cm}^2$  in the relative velocity range of  $\beta \subset [2 \times 10^{-4}, 5 \times 10^{-3}]$ . It can be approximated by a fitting formula uniting analytic, numeric and experimental data [20],

$$\sigma \approx \frac{240a_\infty^2(\beta - \beta_m)^6}{(\beta - \beta_m)^6 + \beta_m^6} \left( \frac{\alpha_F}{\beta + \alpha_F} \right)^2 \ln \left( 1.8 \frac{\beta + \alpha_F}{\alpha_F} \right)$$

where  $a_\infty$  is the Bohr radius,  $\alpha_F$  is the fine structure constant, and  $\beta \geq \beta_m \approx 7.5 \times 10^{-5}$ .

## References

- [1] L.P. Pitaevskii, E.M. Lifshitz, Physical Kinetics, v.10, Pergamon (1981)
- [2] L.D. Landau, JETP, 7, 203 (1937); Phys. Zs. Sowjet, 10, 154 (1936)
- [3] S. Ichimaru, M.N. Rosenbluth, Phys. Fluids, 13, 2778 (1970)
- [4] S. Nagaitsev, PRST-AB 8, 064403 (2005)
- [5] S. Ichimaru, Basic principles of plasma physics (1973)
- [6] H. Boersch, Z. Phys. 139, 115 (1954)
- [7] V. Kudelainen et al, JETP, 56, n.6, 1191 (1982)
- [8] J.D. Bjorken, S.K. Mtingwa, PA 13, p.115 (1983)
- [9] M. Conte, M. Martini, PA 17, 1 (1985)
- [10] A. Piwinski, Proc. 9th Int. Conf. on High Energy Accelerators, Stanford (1974) p.405
- [11] V. Parkhomchuk, Proc. Workshop on Electron Cooling and Related Applications (1984)
- [12] C. Bernardini et al, PRL 10, 407 (1963)
- [13] J. Haissinski, LAL Orsay report 41-63 (1963)
- [14] U. Voelkel, DESY 67/5 (1965)
- [15] Y. Miyahara, Jap. Journal of Appl. Phys., 24, n.9, p.1742 (1985)
- [16] J. Le Duff, CERN Acc. School (1987)
- [17] A. Piwinski, DESY-98-179 (1998)
- [18] V. Lebedev, HB-2004 (AIP Conf. Proc., v.773, 2005) p.440
- [19] M. Chanel et al, PL B192, n.3,4, p.475 (1987)
- [20] V. Lebedev et al, Linac 2010; A. Shishlo et al, PRL (2012)

### 2.4.13 Ion Trapping, Beam-Ion Instabilities, and Dust

F. Zimmermann, CERN

Positively charged ions trapped in the beam potential affect performance of electron and antiproton storage rings. Size of the trapped ions ranges from single atoms, small molecules, to micron-size dust particles. Effects of trapped ions on the beam include: increased emittance, betatron tune shifts and broadening, collective instabilities, reduced beam lifetime. In future storage rings single-pass effects are important, i.e. the ions do not have to be “trapped” to be harmful.

Symbols:  $C$  storage ring circumference,  $f_0$  revolution frequency,  $L_{\text{sep}}$  bunch spacing (in meter), beam distribution assumed bi-Gaussian with rms  $\sigma_{x,y}$ ,  $\epsilon_{x,y}$  unnormalized emittances,  $l_{\text{train}} = n_B L_{\text{sep}}$  bunch-train length,  $b$  beam-pipe radius,  $I$  average beam current,  $\beta_{x,y}$  average  $\beta$ -functions,  $\lambda_{\text{ion}}$  ion line density,  $\sigma_{\text{ion}}$  ionization cross-section,  $\Sigma_{\text{pipe}}$  vacuum pipe cross-sectional area,  $N_{\text{ion}}$  total number of trapped ions ( $N_e$  total number of trapped electrons),  $N = N_B n_B$  total number of particles in beam,  $Q$  ion charge in units of electron charge  $e$  (in general  $Q = 1$  for regular ions,  $Q \gg 1$  for dust particles),  $A$  ion mass in units of proton mass  $m_p$ ,  $A_{\text{atom}}$  and  $Z_{\text{atom}}$  the atomic mass and atomic number of the dust particle constituents,  $f_i = \omega_i/2\pi$  ion oscillation frequency in the beam potential,  $\tau_{\text{th}}$  thermal lifetime of dust particle,  $T_{mp}$  melting point,  $T_{\text{eq}}$  equilibrium temperature of trapped dust particle,  $r_0$  classical radius of the beam particles (in most cases  $r_0 = r_e$  or  $r_p$ ),  $d_{\text{gas}} = p_{\text{gas}}/(k_B T)$  the residual gas density,  $p_{\text{gas}}$  the gas pressure. More symbols are defined in Tabs.1 and 2.

#### 2.4.13.1 Ion trapping

Average beam potential for bi-Gaussian distribution (round pipe, elliptical beam) is [1]

$$U(x, y) \approx -\frac{eN_B}{4\pi\epsilon_0 L_{\text{sep}}} \int_0^\infty dq e^{-\frac{x^2}{2\sigma_x^2+q}-\frac{y^2}{2\sigma_y^2+q}} - e^{-\frac{b^2}{2\sigma_x^2+q}} \times \frac{\sqrt{2\sigma_x^2+q}\sqrt{2\sigma_y^2+q}}{\sqrt{2\sigma_x^2+q}\sqrt{2\sigma_y^2+q}}$$

Table 1: Selected properties and effects of trapped ions.

| property/effect                       | formula   |
|---------------------------------------|---|
| trapping condition (no gap) [4]       | $A \geq A_{\text{crit}} \equiv \frac{QN_B r_p L_{\text{sep}}}{2\sigma_y(\sigma_x + \sigma_y)}$ or $\pi L_{\text{sep}} f_i \leq c$ |
| trapping with gap $T_g$ [1]           | $ \cos(\omega_i l_{\text{train}}/c) - \omega_i T_g \sin(\omega_i l_{\text{train}}/c)/2  < 1$                                      |
| single-ion oscillation frequency [4]  | $f_i = \frac{c}{2\pi} \left( \frac{2QN_B r_p}{AL_{\text{sep}}\sigma_y(\sigma_x + \sigma_y)} \right)^{1/2}$                        |
| equil. ion distr. (no gap) [5]        | $\rho_y = \frac{1}{\pi\sqrt{2\pi}\sigma_y} K_0(\xi) e^{-\xi}$ with $\xi \equiv [y/(2\sigma_y)]^2$                                 |
| coherent tune shift                   | $\Delta\nu_{x,y;\text{coh}} \approx \beta_{x,y} r_0 Q \lambda_{\text{ion}} C / [4\pi\gamma\sigma_{x,y}(\sigma_x + \sigma_y)]$     |
| incoherent tune shift                 | $\Delta\nu_{x,y;\text{inc}} \approx 2\Delta\nu_{x,y;\text{coh}}$  |
| inc. phase shift along train (1 pass) | $\Delta\psi_{x,y}(s) \approx \beta_{x,y} r_p Q \lambda_{\text{ion}} s / [\gamma\sigma_{x,y}(\sigma_x + \sigma_y)]$                |

Table 2: Instability growth rates for classical (multi-turn) ion-related instabilities;  $n$  closest integer,  $\tau_{\text{ion}} = 1/(\sigma_{\text{ion}} c d_{\text{gas}})$ ,  $\nu_p \equiv \sqrt{\frac{2\lambda_e r_p R^2}{\gamma_p \sigma_y(\sigma_x + \sigma_y)}}$ ,  $\nu_e \equiv \sqrt{\frac{N_r e R}{\pi\sigma_y(\sigma_x + \sigma_y)}}$ ,  $\Delta\nu_y \equiv \frac{r_b \beta_y \lambda_{\text{ion}} R}{\gamma\sigma_y(\sigma_x + \sigma_y)}$ ,  $\eta$  desorption yield, and  $s_p$  linear pumping speed (in  $\text{m}^2 \text{s}^{-1}$ ).

| instability   | growth rate $1/\tau$  |
|---|---|
| proton-ion oscillations [9]                                     | $(n - \nu_{x,y})A/(8\nu_{x,y}\gamma\tau_{\text{ion}})$      |
| proton-electron (or antiproton-ion) inst. w/o freq. spread [11] | $\pi f_0 \nu_p \sqrt{\frac{\nu_e}{n - \nu_e}}$              |
| electron-ion & proton-electron inst. w freq. spread [10]        | $\pi^2 f_0 \Delta\nu_y / (\Delta f_i/f_i)_{\text{FWHM}}$    |
| pressure bump [14, 15]  | $(\eta I \sigma_{\text{ion}}/e - s_p)/\Sigma_{\text{pipe}}$ |

$$\approx \frac{IZ_0}{4\pi} \begin{cases} \text{constant} + \frac{x^2}{\sigma_x(\sigma_x + \sigma_y)} + \frac{y^2}{\sigma_y(\sigma_x + \sigma_y)} \\ - \frac{\sigma_x \sigma_y}{(\sigma_x + \sigma_y)^2} \left[ \frac{x^2 y^2}{2\sigma_x^2 \sigma_y^2} + \frac{x^4}{12\sigma_x^4} \left( 1 + \frac{2\sigma_x}{\sigma_y} \right) \right. \\ \left. + \frac{y^4}{12\sigma_y^4} \left( 1 + 2\frac{\sigma_y}{\sigma_x} \right) \right] \\ \text{for } x, y \leq \sigma_x, \sigma_y \\ \text{constant} + \left( \frac{x^2}{\sigma_x(\sigma_x + \sigma_y)} + \sqrt{2} \frac{|y|}{\sigma_y} \right) \\ \text{for } \sigma_y \ll y \ll \sigma_x, x \ll \sigma_x \\ \ln \left( \frac{x^2 + y^2}{b^2} \right) \quad \text{for } x, y \gg \sigma_x, \sigma_y \end{cases} \quad (1)$$

where  $Z_0 = 377 \Omega$ .

Refs.[1, 2, 3, 4] give overview and discussion of ion stability, accumulation, removal and effects on beam; Ref.[1] discusses longitudinal ion motion; Refs.[6, 9, 11, 12, 13] present theories and observations of conventional (i.e. multi-turn) coherent beam-ion/electron instabilities. Tabs.1 and 2 give a summary of these effects.

Techniques for ion removal include (i) clearing gap; (ii) beam shaking (rf knock-out) [6, 7, 8]; (iii) clearing electrodes.

#### 2.4.13.2 Dust particles

Capture of positively charged dust particles has been observed at many electron and antiproton storage rings [16, 17, 18, 19, 20]. To remove or study dust, specific devices have been built, which shake the vacuum chamber [21] or create an electric discharge [22]. Dust trapping in an electron ring has been directly recorded by video camera, showing a luminous body resembling a shooting star [22]. Falling or negatively charged dust particles can also lead to temporary high local losses from proton beams [23].

The estimated *equilibrium temperature*  $T_{\text{eq}}$  of a trapped dust particle allows predicting whether it explodes after being trapped by the beam, or,

in case it survives (when  $T_{\text{eq}} < T_{\text{mp}}$ ), what its thermal lifetime will be.  $T_{\text{eq}}$  is determined by a balance of ionization energy deposition and heat radiation, i.e.

$$\frac{dE}{dt} \Big|_{\text{ion}} = \frac{dE}{dt} \Big|_{\text{rad}} \Rightarrow T_{\text{eq}} \quad (2)$$

Very roughly, for a spherical particle of radius  $R$  ( $R$  is not the ring radius here!),

$$\frac{\Delta E}{\Delta t} \Big|_{\text{ion}} \approx \frac{2R^3 N_B n_B f_0 \rho}{15 \sigma_x \sigma_y} \frac{dE}{d(\rho x)} \Big|_{\text{min}} \quad (3)$$

where  $\rho$  = mass density,  $dE/d(\rho x)|_{\text{min}} \approx 1.5 \text{ MeV cm}^2 \text{ g}^{-1}$ , and

$$\frac{\Delta E}{\Delta t} \Big|_{\text{rad}} \approx - \int_0^\infty d\omega \frac{R^2 \hbar \omega^3 / (\pi c^2)}{e^{\frac{\hbar \omega}{k_B T}} - 1} \bar{Q}(\omega, R) \quad (4)$$

Absorption coefficient  $\bar{Q}(\omega, R)$  for small particles can be found according to Mie theory [19, 24].  $\bar{Q}$  depends on the material and also on shape of the dust particle. It is calculated from the complex dielectric function, e.g., using program BH-MIE [24]. If  $T_{\text{eq}} > T_{\text{mp}}$ , particle explodes. If  $T_{\text{eq}} < T_{\text{mp}}$ , the thermal lifetime due to evaporation is

$$\frac{1}{\tau_{\text{th}}} \equiv \frac{\dot{A}}{A} \approx - \frac{4\pi A_{\text{atom}}}{\sqrt{2\pi m_p A_{\text{atom}} k_B}} \times \left( \frac{3m_p}{4\pi\rho} \right)^{2/3} \frac{p(T_{\text{eq}}(A)) [\text{Pa}]}{\sqrt{T_{\text{eq}}(A)} A^{1/3}} \quad (5)$$

where  $p(T)$  is the vapor pressure as a function of temperature. Eqs.(2-5) are solved numerically. Illustrative example shown in Fig.1 [19].

The dynamic stability of a trapped dust particle, as well as the betatron tune shift and the

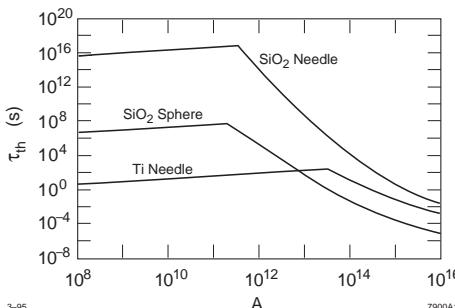


Figure 1: Thermal lifetime  $\tau_{\text{th}}$  of trapped dust particles in the HERA electron ring as a function of mass.

beam-lifetime reduction that such a particle may induce, are determined by its *equilibrium charge*. The equilibrium charge is calculated by equating the ionization rate

$$\dot{Q}_{\text{ioniz}} \approx \frac{4\pi}{3} f_0 N_A N_A r_e \frac{Z_{\text{atom}}}{A_{\text{atom}}} \frac{r_e}{\sigma_x \sigma_y} \frac{R^4}{Q} \quad (6)$$

and the field evaporation of ions [25]

$$\dot{Q}_{\text{ev}} \approx - \frac{A_{\text{atom}} m_p 8\pi^2 k_B T^2}{h^3} \times \exp \left[ - \frac{U + V - \Phi_-}{k_B T} + \frac{e^2 \sqrt{Q}}{4\pi \epsilon_0 R k_B T} - \frac{1}{k_B} \int_0^T \frac{dT'}{T'^2} \int_0^{T'} dT'' C_p(T'') \right] \quad (7)$$

where  $C_p$  is the heat capacity at constant pressure,  $N_A$  is Avogadro's number,  $U$  and  $V$  ionization and vaporization energy, respectively, and  $\Phi_-$  workfunction. The trapping of photoelectrons is another discharging mechanism.

Finally, there are two effects by which a trapped dust particle can reduce the *beam lifetime*:

(i) Bremsstrahlung on nuclei field [26]

$$\frac{1}{\tau_b^n} \approx - \frac{c A Z_{\text{atom}}^2}{A_{\text{atom}} \sigma_x \sigma_y C} \left( \frac{8r_e^2 \alpha}{3\pi} \ln \frac{\Delta E}{E} \ln \frac{183}{Z_{\text{atom}}^{1/3}} \right)$$

where  $\Delta E/E$  is energy acceptance;

(ii) Bremsstrahlung in collective field of dust particle [19]

For needle-shaped particle of transverse radius  $R$  and length  $h$ , aligned in beam direction,

$$\frac{1}{\tau_b^{\text{coll}}} \approx \pi \int_0^{r_{\text{max}}} N_\gamma(r) r dr \frac{c}{2\pi \sigma_x \sigma_y C} \quad (8)$$

$$N_\gamma(r) \approx \frac{5}{2\sqrt{3}} \frac{c \alpha \Upsilon(r)}{\lambda_e \gamma} \frac{\max(r, h)}{[1 + \Upsilon(r)^2/3]^{1/2}}$$

$$\Upsilon(r) \approx \frac{2\hbar \gamma Q r_e}{c} \frac{\min(r/h, 1)}{\max(R, r)^2}$$

where  $r_{\text{max}}$  is defined by  $\Upsilon(r_{\text{max}}) = 0.02$  [27]. In most cases, this second effect is not important.

#### 2.4.13.3 Single-pass ion effects in storage rings and linacs

*Nonlinear coupling* effect due to ions can cause vertical emittance growth. It is contained if betatron phases have  $|\mu_x - \mu_y|/|\mu_x| > 5\%$  [28]. However, *resonant single-bunch tail excitation* can still occur (Fig.2). An initial perturbation  $y_0(z, s) =$

## Sec.2.4: COLLECTIVE EFFECTS

Table 3: Rise-time estimates for fast beam-ion instability and equilibrium with feedback and noise;  $\tilde{\omega}_i \equiv c \left( \frac{4N_B r_p}{3AL_{\text{sep}} \sigma_y (\sigma_x + \sigma_y)} \right)^{1/2}$  the coherent angular ion frequency ( $Q = 1$  assumed),  $(\Delta\tilde{\omega}_i)_{\text{rms}}$  its rms variation around the ring circumference, and  $\bar{\eta}(z) \equiv z\sqrt{\tau_d/\tau_c}/(n_B L_{\text{sep}})$ , where  $z$  denotes the longitudinal position along the bunch train ( $z = 0$  at the start of the train),  $\tau_d$  is the (feedback) damping time, and  $\langle y^2(0) \rangle$  the rms oscillation amplitude of the first bunch in the train.

| approximation                                  | asymptotic behavior   | characteristic rate $1/\tau$  |
|--|---|---|
| linear treatment ( $y \ll \sigma_y$ ) [30]     | $y \sim \exp(\sqrt{t/\tau_c})$  | $\frac{1}{\tau_c} \equiv \frac{4d_{\text{gas}}\sigma_{\text{ion}}\beta_y N_B^{3/2} n_B^2 r_e r_p^{1/2} L_{\text{sep}}^{1/2} c}{3\sqrt{3}\gamma\sigma_y^{3/2}(\sigma_x + \sigma_y)^{3/2} A^{1/2}}$ |
| with decoherence & spread $\Delta f_i$ [32]    | $y \sim \exp(t/\tau_e)$   | $\frac{1}{\tau_e} \approx \frac{1}{\tau_c} \frac{c}{2\sqrt{2}l_{\text{train}}(\Delta\tilde{\omega}_i)_{\text{rms}}}$  |
| nonlinear regime ( $y \gg \sigma_y$ ) [32]     | $y \sim t/\tau_H$   | $\frac{1}{\tau_H} \approx \frac{1}{\tau_c} \frac{c}{\tilde{\omega}_i L_{\text{sep}} n_B^{3/2}}$   |
| equilibrium state with feedback and noise [33] | $\langle y^2(z) \rangle \approx \sqrt{\tau_d/\tau_c} \langle y^2(0) \rangle \bar{\eta}^3/(24 n_B)$ (for $\bar{\eta} \ll 1$ )<br>$\langle y^2(z) \rangle \approx \sqrt{\tau_d/\tau_c} \langle y^2(0) \rangle e^{\bar{\eta}^2/2}/(4\bar{\eta} n_B)$ (for $\bar{\eta} \gg 1$ ) |   |

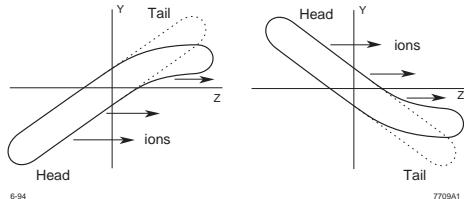


Figure 2: Schematic picture of resonant tail excitation for two different betatron phases; beam moves to the left.

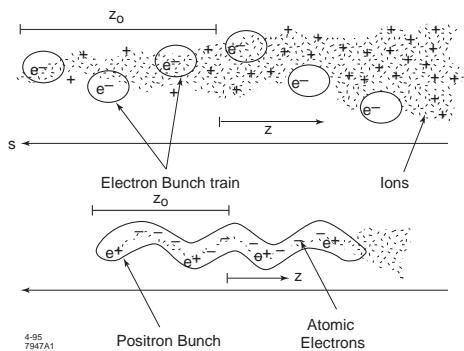


Figure 3: Schematic of fast beam-ion instability which can arise due to ion trapping in electron bunch train or due to trapping of free electrons in single positron bunch.

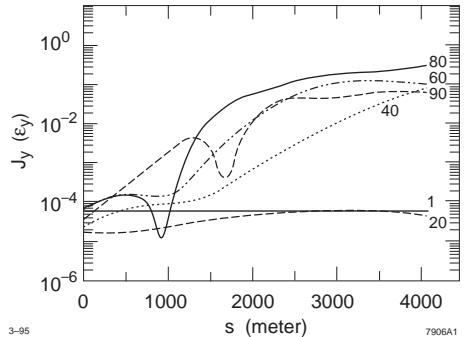


Figure 4: Simulated growth of the action of the vertical centroid for every twentieth bunch in the NLC Damping Ring for a vacuum of  $10^{-8}$  Torr of CO gas; growth saturates at  $\sim \sigma_y$  due to nonlinearity of beam-ion force.

$\hat{y}z/\sigma_z \cos(s/\beta_y + \phi_0)$ , e.g., due to wakefields, leads to emittance growth after distance  $L$  [29],

$$\Delta\epsilon_y(L) \approx \frac{1}{2\pi\beta_y} \left( \frac{\hat{y}L\beta_y\lambda_{\text{ion}}r_e}{2\gamma\sigma_y(\sigma_x + \sigma_y)} \right)^2 \quad (9)$$

where  $\lambda_{\text{ion}}$  is the ion density at end of bunch.

Figures 3 and 4 illustrate the *fast beam-ion instability* mechanism. The theory and simulations, are discussed in [30]-[35] and summarized in Tab.3. (Possible) experimental evidence is discussed in [36]-[38].

## References

- [1] D. Villevaud, S. Heifets, PEP-II AP-18-93 (1993)
- [2] A. Poncet, Lecture Notes in Phys. 425, Springer Verlag (1994) p.202
- [3] Y. Baconnier, A. Poncet, P. Tavares, CERN/PS 94-40 (1994)
- [4] Y. Baconnier, G. Brianti, CERN/SPS/80-2 (1980)
- [5] T.O. Raubenheimer, F.-J. Decker, J.T. Seeman, PAC 95, p.3291
- [6] R. Alves Pires et al, PAC 89, p.800
- [7] J.C. Lee et al, PAC 95, p.2968
- [8] S.J. Werkema, PAC 95, p.3397
- [9] H.G. Hereward, CERN MPS/Int. DL 64-8 (1964)
- [10] H.G. Hereward, CERN 71-15 (1971)
- [11] E. Keil, B. Zoller, CERN-ISR-TH/71-58 (1971)
- [12] G. Koshkarev, P. Zenkevich, PA 3 (1972) 1
- [13] L.J. Laslett, A.M. Sessler, D. Möhl, NIM A121 (1974) 517
- [14] E. Fischer, K. Zankel, CERN/ISR-VA/73-52(1973)
- [15] O. Gröbner, R.S. Calder, PAC 73, p.760
- [16] H. Saeki et al, RSI 62, No.4, 874 and No.11 (1991) 2558
- [17] D. Sagan, NIM A330 (1993) 371
- [18] E. Jones et al, PAC 85, p.2218
- [19] F. Zimmermann et al, PAC 95, p.517
- [20] D.R.C. Kelly et al, PAC 95, p.2017
- [21] H. Frischholz, T. Wikberg, CM-P00072614, CERN/ISR-VA/TW/ck (1971)
- [22] Y. Tanimoto et al, PRST-AB 12, 110702 (2009)
- [23] T. Baer et al; F. Zimmermann et al, IPAC'11
- [24] C.F. Bohren, D.R. Huffman, Absorption and Scattering of Light by Small Particles, Wiley (1983)
- [25] E.U. Condon, H. Odishaw, Handbook of Physics, 2nd ed., McGraw-Hill (1967)
- [26] A. Piwinski, CERN 85-19 (1985)
- [27] K. Yokoya, P. Chen, Lect. Notes in Phys. 400, Springer Verlag (1990) p.415
- [28] T. Raubenheimer, P. Chen, LINAC 92
- [29] P. Emma, T. Raubenheimer, F. Zimmermann, EPAC 94, p.1162
- [30] T.O. Raubenheimer, F. Zimmermann, PR E52, No.5 (1995) 5487
- [31] G.V. Stupakov et al, PR E52, 5 (1995) 5499
- [32] G.V. Stupakov, Proc. Int. Workshop on Collective Effects and Impedance for B factories, KEK Proc. 96-6 (1996) p.243
- [33] A.W. Chao, G.V. Stupakov, Proc. MBI97 workshop (1997)
- [34] G.V. Stupakov, PRST-AB 3 (2000) 019401
- [35] R. A. Bosch, PRST-AB 3 (2000) 034402
- [36] J. Byrd et al, PRL, Vol.79, No.1 (1997) 79
- [37] J. Huang et al, PRL 81 (1998) 4388
- [38] Y. Ohnishi et al, EPAC 00, p.1167

### 2.4.14 Electron-Cloud Effect

*M.A. Furman, LBNL & Cornell U.*

The electron-cloud effect (ECE) was first identified at the Photon Factory (PF) at KEK [1] as a fast transverse coupled-bunch instability that arose only when PF was operated with a positron  $e^+$  beam. Unlike the ion-induced instability (Sec.2.4.13) observed when operated with an electron beam, the  $e^+$  beam instability persisted even with a substantial gap in the bunch train. The coupled-bunch mode spectrum for the  $e^+$  beam was qualitatively different from that for an  $e^-$  beam under otherwise similar conditions. The phenomenon disappeared when the bunch spacing was sufficiently large, and it could not be attributed to known machine impedances. The amplitude of the unstable motion reached saturation and was accompanied by the excitation of vertical coupled-bunch oscillations, and possibly of vertical emittance growth.

Experimental analysis [1], simulations [2] and analytical work [3] showed that the cause of the instability is an electron cloud (EC) that developed inside the vacuum chamber generated by photoelectron emission by synchrotron radiation from the beam striking the walls of the chamber. This photoelectron instability (PEI) [2] is one of many ECEs investigated in positron storage rings since then. The phenomenon was later studied in dedicated experiments at BEPC and APS [4, 5]. The ECE led to serious performance limitations at PEP-II and KEKB [6]. A closely related coupled-bunch instability was previously observed at CESR, although in this case the photoelectrons were trapped and localized in a section of the ring rather than spread out over most of the circumference. A comprehensive program dedicated to measurements and analysis of ECE's for  $e^+e^-$  storage rings is now ongoing at CESR [7].

The above-mentioned ECE's are related to previously observed electron-proton dynamical effects such as beam-induced multipacting (BIM), first observed at ISR [8] when operated with bunched beams. Closely related to BIM is trailing-edge multipacting observed at the LANL spallation neutron source PSR [9], where electron detectors register a large signal during the passage of the tail of the bunch even for stable beams.

All ECEs in  $e^+e^-$  as well as in hadron storage rings have precursors in the e-p instabilities for bunched and unbunched beams first seen at BINP in the mid-60s [10].

**Phenomenology** In  $e^+$  or  $e^-$  rings, the EC typically starts when the synchrotron radiation (SR) emitted by the beam yields photoelectrons upon striking the vacuum chamber. In proton rings, the EC is typically initiated by ionization of residual gas, or from electron generation when stray beam particles strike the chamber. The LHC is the first proton storage ring in which the beam emits copious SR (critical energy  $\sim 44$  eV), leading to substantial photoemission, hence the mechanism for the formation of the EC is analogous to present-day positron rings [11].

The above-mentioned primary mechanisms are typically insufficient to lead to a significant EC density  $n_e$ . However, these primary electrons are kicked by successive bunches, striking the chamber walls with a broad energy spectrum and leading to secondary electron emission (SEE). This can lead to a large amplification ( $\sim$  a few orders of magnitude) of the primary electron density and to strong temporal and spatial fluctuations in the electron distribution [12]. This compounding effect of SEE is usually the main determinant of the strength of the ECEs, and is particularly strong in positively-charged bunched beams.

The ECE combines many parameters of a storage ring such as bunch current, bunch spacing, beam energy [13], vacuum chamber geometry, vacuum pressure, and electronic properties of the chamber surface material such as photon reflectivity  $R_\gamma$ , effective photoelectric yield (or quantum efficiency)  $Y_{\text{eff}}$ , secondary electron yield (SEY), characterized by the function  $\delta(E)$  ( $E$  = electron-wall impact energy), secondary emission spectrum [14, 15], etc. The function  $\delta(E)$  has a peak  $\delta_{\max}$  typically ranging in 1 – 4 at an energy  $E = E_{\max}$  typically ranging in 200 – 400 eV. A convenient phenomenological parameter is the effective SEY,  $\delta_{\text{eff}}$ , defined to be the average of  $\delta(E)$  over all electron-wall collisions during a relevant time window. Unfortunately, there is no simple a-priori way to determine  $\delta_{\text{eff}}$ , because it depends in a complicated way on a combination of many of the above-mentioned beam and chamber parameters. If  $\delta_{\text{eff}} < 1$ , the chamber walls act as net absorbers of electrons and  $n_e$  grows linearly in time following beam injection into an empty chamber. The growth saturates

when the net number of electrons generated by primary mechanisms balances the net number of electrons absorbed by the walls. If  $\delta_{\text{eff}} > 1$ , the EC grows exponentially. The exponential growth slows down as the space-charge fields from the EC effectively neutralize the beam field, reducing the electron acceleration. Ultimately, the process stops when the EC space-charge fields are strong enough to repel the electrons back to the walls of the chamber upon being born, at which point  $\delta_{\text{eff}}$  becomes = 1. At this point, the EC distribution reaches a dynamical equilibrium characterized by rapid temporal and spatial fluctuations, determined by the bunch size and other variables. For typical present-day storage rings, whether using  $e^+$  or  $p$  beams, the spatio-temporal average  $n_e$  reaches a level  $\sim 10^{10-12} \text{ m}^{-3}$ . The energy spectrum of the electrons typically peaks at an energy below  $\sim 100$  eV, and has a high-energy tail reaching out to keV's. If there is a gap in the beam, or if the beam is extracted, the cloud dissipates with a falltime that is controlled by the low-energy value of  $\delta(E)$  [16]. In general, there is no simple, direct correlation between the risetime and the falltime. Electron generation, trapping and ejection from quadrupole magnets at the PSR is significant, and these electrons act as seeds for the EC buildup in nearby drift regions [17].

In regions of the storage ring with an external magnetic field, such as dipoles, quadrupoles, etc., the EC distribution develops characteristic geometrical patterns. For typical B-fields in the range 0.01–5 T and typical EC energies  $< 100$  eV, the electrons move in tightly-wound spiral trajectories about the field lines. In practice, in a bending dipole, the electrons are free to move in the vertical ( $y$ ) direction, but are essentially frozen in the horizontal ( $x$ ). As a result, the  $y$ -kick imparted by the beam on a given electron has an  $x$  dependence that is “remembered” by the electron for many bunch passages. It often happens that the electron-wall impact energy equals  $E_{\max}$  at an  $x$ -location less than the chamber radius. At this location  $\delta(E) = \delta_{\max}$ , hence  $n_e$  is maximum, leading to characteristic high-density vertical stripes symmetrically located about  $x = 0$  [18]. For quadrupole magnets, the EC distribution develops a characteristic four-fold pattern, with characteristic four-fold stripes [19].

**Diagnostics** Dedicated electron detectors have been designed and used to study the intensity and energy spectrum of the electrons striking the

vacuum chamber. The EC average density can be inferred via simulations, in some cases, from the measured coherent tune shift of successive bunches along a train. Other techniques, notably the use of microwaves injected into the vacuum chamber, have been employed. See Sec.7.4.15.

**Consequences** One consequence of the EC is a strong increase in the vacuum pressure as a function of bunch current. The pressure rise exhibits a threshold behavior, and is sensitive to the bunch fill pattern at fixed total current, as observed in the B factories and the SPS [20]. When RHIC was operated with ion beams, the pressure rise was rather dramatic at transition energy due to the short bunch length, often triggering a beam abort by the machine protection system [21] (the problem has been controlled via low-SEY coatings—see below).

In 1996-97, calculations showed that the LHC will be subject to an ECE [11, 22, 23]. The main concern here is the power deposited by the electrons rattling around the chamber, which must be dissipated by the cryogenic system. Since the cryogenic system was designed before the discovery of the ECE, substantial effort has been devoted since 1997 to better estimate the power deposition, to identify the conditions under which the cooling capacity may be exceeded, and to devise mitigation mechanisms if necessary. As part of this effort, the ECE has been studied at SPS and PS at the high beam intensities required for nominal LHC operation. A comprehensive reference database can be found in [24].

EC-induced single-bunch head-tail instabilities [25, 26, 27] have been predicted and observed at several storage rings. Before EC mitigation mechanisms were implemented at KEKB, such instabilities led to an effective beam blowup, limiting luminosity performance. Bunch-to-bunch tune shift that grows towards the tail of the bunch train, sometimes leading to coupled-bunch instabilities, were observed [1, 28, 29].

Electrons from the cloud getting “sucked into” the body of a passing bunch typically lead to a rather significant head-tail tune spread. When this tune spread is combined with synchrotron oscillation and/or space-charge forces, incoherent effects result, such as slow emittance growth and/or particle losses. The bunch current threshold for such effects can be lower than for the above-mentioned single-bunch instabilities [30, 31].

A high-current instability that has been observed for many years at the PSR is also an ECE [9]. The phenomenon has been studied in intense, long-pulse, heavy-ion fusion drivers [32], at the J-PARC proton rings [33], and at the FNAL Main Injector [34].

If a train of bunches, separated by a time interval  $\tau$ , traverses a region with a dipole magnetic field  $B$ , beam-EC resonances leading to an enhancement of  $n_e$  occur when  $\tau$  is an integer multiple of the electron cyclotron period, i.e. when  $\omega\tau/2\pi = \text{integer}$ , where  $\omega = eB/m$ ,  $m$  being the electron mass [35]. The strength of these resonances is characterized by the phase-averaging factor  $e^{-\omega^2\sigma^2/2}$ , where  $\sigma = \text{rms bunch length (in units of time)}$ . Consequently, they are well defined only for short bunches, i.e., for  $\omega\sigma \lesssim 1$ .

**Mitigating mechanisms** A low value of  $Y_{\text{eff}}$  is favorable, as  $n_e$  is a monotonically increasing function of  $Y_{\text{eff}}$ . At the LHC small transverse grooves were engraved on the vacuum chamber, effectively resulting in normal photon incidence thus leading to a reduction of  $Y_{\text{eff}}$  by a factor of 2–4, and  $R_\gamma$  by a factor  $\sim 10$  [36].

$Y_{\text{eff}}$  can also be reduced by an antechamber on the outboard side of the vacuum chamber. At PEP-II [37], the antechamber allows for  $\sim 99\%$  of the photons to escape out of the vacuum chamber. The remaining 1% of the photons, however, are radiated at wide angle and low energy, generating photoelectrons more efficiently [38] than the high-energy photons that escape.

In practice, lowering  $Y_{\text{eff}}$  is effective only when SEE is not dominant. If SEE off the chamber surface is significant, achieving a low value of  $\delta_{\text{eff}}$  is essential. Lowering  $Y_{\text{eff}}$  via antechambers is a practical option only when implementable, as is often the case in Al chambers made by extrusion. Although pure Al has a low peak value of  $\delta$ ,  $\delta_{\text{max}} \lesssim 1$ , its surface is normally covered with a layer of  $\text{Al}_2\text{O}_3$  with  $\delta_{\text{max}} \sim 2.5 - 3$ , among the highest of all metals used in practice to build vacuum chambers. Copper and stainless steel have  $\delta_{\text{max}} \sim 1.3 - 1.5$  when adequately conditioned, but this value may not be low enough in many cases. Beryllium, used in rf cavity windows and in the vicinity of IP of a collider, has similar features to  $\text{Al}_2\text{O}_3$ : its oxide has  $\delta_{\text{max}} \sim 4 - 5$ , but its elemental form has  $\delta_{\text{max}} \sim 0.6 - 0.9$ .

The ECE is a self-conditioning effect in the sense that  $\delta_{\text{max}}$  gradually decreases owing to the electron-wall bombardment during normal

machine operation [9]. The important practical question is how long it takes for  $\delta_{\max}$  to fall below a value where the EC is no longer an operational limitation.

Low-emission coatings are usually employed to reduce SEE. For example, the Al chambers in the arcs of the PEP-II e<sup>+</sup> ring were coated with TiN [39]. This coating, once properly conditioned, has  $\delta_{\max} \sim 1.1$ . TiN coatings are used in the SNS [40] and are recommended for the damping rings of a future linear collider [41]. Other coatings, such as the NEG substance TiZrV, have been shown to reduce SEE and have been, or will be, used in the warm regions of RHIC and the LHC [42, 43], and possibly in the ILC damping rings [44]. Amorphous (glassy) carbon is being actively investigated as an effective coating for the LHC [45, 46].

Longitudinal grooves (parallel to the beam direction) of pitch and depth on the mm scale etched on the vacuum chamber walls have shown to suppress SEY [47].

Reducing  $\delta_{\max}$ , however, is sometimes not enough: the low-energy ( $E \lesssim 10$  eV) value of  $\delta(E)$ , and certain details of the emission energy spectrum, primarily related to the intensity of the redifused component [15, 48, 14], can have a significant effect on the survival of the EC during a beam gap and hence on the average  $n_e$ .

Active mechanisms have been used to control EC-induced instabilities, such as raising the vertical chromaticity above its nominal value, or using octupoles [49, 50]. Solenoidal windings wrapped around most of the circumference of the e<sup>+</sup> rings of PEP-II and KEKB, proved essential for these machines to reach, and exceed, their specified performance [6, 51]. Solenoids have also been used in parts of the circumference of RHIC [21]. A relatively low solenoidal field ( $B \sim 20 - 30$  G) is sufficient to trap the electrons near the walls of the vacuum chamber, thereby minimizing their effect on the beam. Elaborate bunch fill patterns were used at PEP-II [52] and RHIC [21]. The many gaps in these patterns have the effect of promoting the dissipation of the EC. At PSR it was found that increasing the momentum spread of the beam increases the current instability threshold [9]. Clearing electrodes have been designed and tested at KEKB [53]. To the extent that the EC-induced instability has some degree of coherence, it is in principle possible to combat it via a feedback system. High-bandwidth (up to  $\sim 1$  GHz) systems

are being studied to control the single-bunch EC-induced instability observed at the SPS [54].

Besides the usual accelerator conference proceedings and internet websites of the ECLOUD and HB workshop series, other useful websites can be found at:

- Pressure Rise 2003 (BNL):

- <http://www.c-ad.bnl.gov/icfa/>

- Stopping & range of electrons in matter:

- <http://www.srim.org/SREM.htm>

- UT-Knoxville metrology & lithography:

- <http://web.utk.edu/~srcutk/>

- EC in the LHC:

- <http://ab-abp-rlc.web.cern.ch/ab-abp-rlc-ecloud/>

- Comparison of EC Simulations:

- <http://wwwslap.cern.ch/collective/ecloud02/ec-sim/index.html>

- CERN website on SEE:

- [http://at-div-vac.web.cern.ch/at-div-vac/VACPAGES/ps/Phys&tech/Phys/Ecloud/SEY/SEY\\_paper.html](http://at-div-vac.web.cern.ch/at-div-vac/VACPAGES/ps/Phys&tech/Phys/Ecloud/SEY/SEY_paper.html)

- NIST database on electron impact cross-sections:

- <http://physics.nist.gov/PhysRefData/Ionization/>

- LBNL X-ray interactions with matter:

- [http://henke.lbl.gov/optical\\_constants/](http://henke.lbl.gov/optical_constants/)

## References

- [1] M. Izawa et al, PRL 74, 5044 (1995)
- [2] K. Ohmi, PRL 75, 1526 (1995)
- [3] S. Heifets, Proc. CEIBA95, p.295
- [4] Z.Y. Guo et al, PAC97, 3B012
- [5] K.C. Harkay, PAC99, TUAL4
- [6] H. Fukuma, ICFA BDNL 48, p.112 (Apr. 2009)
- [7] M.A. Palmer et al, PAC09, FR1RAI02
- [8] O. Gröbner, 10th Intl. Conf. on Part. Accel, 1977, p.277
- [9] R.J. Macek et al, ECLOUD04, p.93
- [10] V. Dudnikov, PAC2001, TPPH094
- [11] F. Zimmermann, SLAC-PUB-7425, LHC Project Report 95 (1997)
- [12] G. Rumolo et al, PRST-AB 4, 012801 (2001); Erratum: 4, 029901 (2001)
- [13] G. Rumolo et al, PRL 100, 144801 (2008)
- [14] M.A. Furman, LBNL-60512/CBP Note-736 (2006); Proc. HB2006, <http://hb2006.kek.jp>
- [15] R. Cimino et al, PRL 93, 014801 (2004)
- [16] M.A. Furman, PAC03, TOPC001
- [17] R. Macek et al, ECLOUD10
- [18] F. Zimmermann, ECLOUD02, p.47
- [19] G. Arduini et al, EPAC04, WEPLT044
- [20] G. Arduini et al, ECLOUD04, p.31; J.M. Jiménez, p.49

- [21] W. Fischer et al, PRST-AB 11, 041002 (2008)
- [22] O. Gröbner, PAC97, p.3589
- [23] M.A. Furman, LBNL-41482/CBP Note 247/CERN LHC Report No. 180 (1998)
- [24] EC in the LHC, see URL list above.
- [25] K. Ohmi, F. Zimmermann, PRL 85(18), 3821 (2000)
- [26] K. Ohmi et al, PRE 65, 016502 (2002)
- [27] F. Zimmermann, PRST-AB 7, 124801 (2004)
- [28] J.A. Crittenden et al, IPAC10, TUPD024
- [29] K. Ohmi et al, ECLOUD04, p.313
- [30] G. Franchetti et al, PRST-AB 12, 124401 (2009)
- [31] J.-L. Vay et al, IPAC10, WEOBRA02
- [32] A.W. Molvik et al, POP 14, 056701 (2007)
- [33] T. Toyama et al, ECLOUD04, p.77
- [34] K.G. Sonnad et al, PAC07, FRPMS028
- [35] C.M. Celata et al, PRST-AB 11, 091002 (2008)
- [36] V. Baglini et al, EPAC 98, TUP18H
- [37] M.A. Furman et al, MBI97, p.170
- [38] O. Gröbner et al, J. Vac. Sci. Technol. A7, 223 (1989)
- [39] K. Kennedy et al, PAC97, 8C009
- [40] P. He et al, PAC01, p.2159
- [41] M. Pivi et al, IPAC10, WEPE097
- [42] LHC Design Report, CERN-2004-003, 1, p.346
- [43] Y. Suetsugu et al, PAC07, FRPMN042
- [44] M. Pivi et al, PAC05, ROPB001
- [45] E.N. Shaposhnikova et al, PAC09, MO6RFP008
- [46] C. Yin Vallgren et al, IPAC11, TUPS027 and TUPS 028
- [47] J.S.T. Ng et al, ICFA BDNL 48 (2009) p.102; Y. Suetsugu et al, p.118
- [48] M.A. Furman, V.H. Chaplin, PRST-AB 9, 034403 (2006)
- [49] Z.Y. Guo et al, MBI97, p.150
- [50] H. Fukuma, Two-Stream2000
- [51] J.T. Seeman, EPAC08, TUXG01
- [52] F.-J. Decker et al, PAC01 TPPH126
- [53] Y. Suetsugu et al, IPAC10, WEOAMH01
- [54] J.D. Fox et al, IPAC10, WEPEB052

## 2.4.15 Coherent Synchrotron Radiation Instability

*G. Stupakov, SLAC*

**Radiation reaction force—CSR wake field**  
The collective force acting on a beam due to its coherent synchrotron radiation (CSR) is described in terms of the so called CSR longitudinal wake [1, 2, 3]. For an ultrarelativistic particle and a long magnet, the plot of the wake (Sec.3.2.1) in the vicinity of a point charge as a function of the distance  $z$  along the particle's orbit is shown in Fig.1, where  $\rho$  denotes the bending radius. The

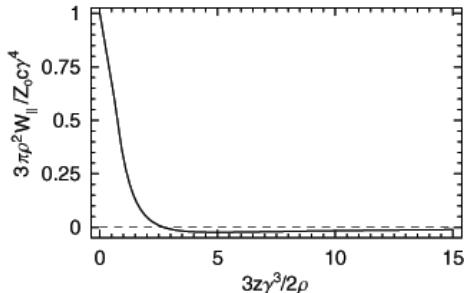


Figure 1: CSR wake  $W_{\parallel}$  as a function of distance  $z$ . The wake is normalized by  $Z_0 c \gamma^4 / (3 \pi \rho^2)$  and the distance is measured in units  $2\rho / (3\gamma^3)$ . The wake vanishes for  $z < 0$ .

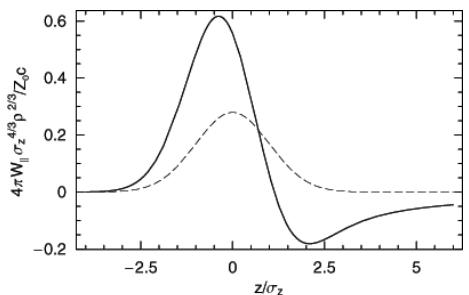


Figure 2: CSR wake (per unit length, solid curve) for a Gaussian bunch (dashed curve) with rms length  $\sigma_z$ . The wake is normalized by  $Z_0 c / (4\pi \sigma_z^{4/3} \rho^{2/3})$  and the distance is measured in units  $\sigma_z$ . The head of the bunch corresponds to  $z > 0$ .

coordinate  $z$  is measured in the direction of motion, and the wake is localized in front of the particle in contrast to “traditional” wakes which trail the source charge. This is because the charge follows a circular orbit and the radiation propagates along chords getting ahead of the source. The wake has the property  $\int_0^\infty W_{\parallel}(z) dz = 0$ . For distances such that  $\rho \gg z \gg \rho/\gamma^3$  (corresponding to abscissa values  $\gtrsim 5$  in Fig.1) the wake (per unit length of path) is

$$W_{\parallel}(z) = -\frac{Z_0 c}{2\pi} \frac{1}{\rho^{4/3} z^{4/3}} \quad (1)$$

The CSR wake convoluted with a Gaussian bunch is shown in Fig.2. The negative value of this wake in the head of the bunch corresponds to energy gain. From Eq.(1) one finds the CSR longitudinal

impedance

$$\begin{aligned} Z_{\parallel}(\omega) &= \frac{1}{c} \int_0^{\infty} dz W_{\parallel}(z) e^{-i\omega z/c} \\ &= \frac{Z_0}{2\pi} \frac{e^{i\pi/6}}{3^{1/3}} \Gamma\left(\frac{2}{3}\right) \frac{\omega^{1/3}}{c^{1/3} \rho^{2/3}} \end{aligned} \quad (2)$$

where  $\Gamma$  is the gamma function. The real part of this impedance is related to the spectrum of the energy loss of a charge due to the radiation:  $dP/d\omega = (e^2/4\pi^2\epsilon_0)\text{Re}Z_{\parallel}$ , where  $dP/d\omega$  is the power of synchrotron radiation per unit length in unit spectral interval. Eq.(2) is valid for  $c/\rho \ll \omega \ll c\gamma^3/\rho$ .

**Wall shielding and formation length** The wake Fig.2 is valid for long magnets, when the magnet length  $l_m$  is much larger than the so called *formation length* of the radiation  $(24\sigma_z\rho^2)^{1/3}$  for a given bunch length. In the opposite limit the wake is dominated by transient effects at the entrance and the exit from the magnet [4].

Close proximity of metallic walls to the beam orbit suppresses the CSR wake. This *wall shielding* effect becomes important when the distance from the beam orbit to the walls  $h$  is such that  $h \lesssim (\sigma_z^2\rho)^{1/3}$ .

**Curvature induced transverse electromagnetic forces** In addition to the longitudinal force there is also a transverse electromagnetic force  $F_{\perp}$  caused by the curvature of the beam trajectory. This force varies in radial direction within the bunch however the gradient  $\partial F_{\perp}/\partial r$  does not cause additional beam focusing to the lowest order in curvature [5]. This effect is referred to as the *cancellation* effect, and is due to the compensation of the gradient of  $F_{\perp}$  by the particle's energy variation in the potential field of the beam space charge [6].

**Instability** The CSR wakefield can cause a microwave instability of a beam in a ring. Theoretical analysis of the instability based on the assumption of short wavelength of the perturbation,  $k\sigma_z \gg 1$ , and the free space CSR wakefield (1) leads to the following conclusions [7]. The beam is unstable for wavenumbers  $k < 2.0\Lambda^{3/2}/\rho$  with the maximum growth rate at  $k = 0.68\Lambda^{3/2}/\rho$  equal to  $(\text{Im } \omega)_{\max} = 0.43\Lambda^{3/2}c\eta\delta_0/\rho$ . The dimensionless parameter  $\Lambda$  is defined by

$$\Lambda = \frac{1}{\eta\gamma\sigma_{\delta}^2} \frac{I}{I_A} \frac{\rho}{\langle \rho \rangle} \quad (3)$$

where  $I$  is the peak beam current,  $I_A = 17.5$  kA,  $\sigma_{\delta}$  is the rms relative energy spread,  $\eta$  is the momentum compaction factor (assumed positive), and  $\langle \rho \rangle = C/2\pi$  with  $C$  the ring circumference. Wall shielding suppresses the instability at long wavelengths and limits the region of instability to  $k \gtrsim \rho^{1/2}/h^{3/2}$ . Further theoretical studies of the CSR instability included detailed computer simulations of the instability in a nonlinear "bursting" phase [8].

A comparison of the theory with observations has been carried out in the experiment at the ALS [9]. The observed thresholds for the CSR bursts agreed well with the instability thresholds from the coasting beam model.

**CSR instability in bunch compressors** Microbunching due to the CSR induced instability has been identified as a potential danger in bunch compressors. Experimental observations [10] showed an increased energy spread and horizontal emittance growth of the beam caused by the coherent radiation. The effect has been studied numerically [11], and several theoretical models of the instability in bunch compressors were developed [14, 12, 13].

## References

- [1] L.V. Iogansen, M.S. Rabinovich, Sov. Phys. JETP, 37 (1960) 83
- [2] Y.S. Derbenev et al, DESY FEL Report TESLA-FEL 95-05 (1995)
- [3] J.B. Murphy, S. Krinsky, R.L. Gluckstern, PAC 95
- [4] E.L. Saldin, E.A. Schneidmiller, M.V. Yurkov, NIM A398 (1997) 373
- [5] E.P. Lee, PA 25 (1990) 241
- [6] Y.S. Derbenev, V.D. Shiltsev, SLAC-PUB-7181 (1996)
- [7] G. Stupakov, S. Heifets PRST-AB 5 (2002) 054402
- [8] M. Venturini, R. Warnock, PRL 89 (2002) 224802
- [9] J. Byrd et al, PRL 89 (2002) 224801
- [10] H. Braun et al, PRST-AB 3 (2000) 124402
- [11] M. Borland, PRST-AB 4 (2001) 070701
- [12] E.L. Saldin, E.A. Schneidmiller, M.V. Yurkov, NIM A490 (2002) 1
- [13] Z. Huang, K.-J. Kim, PRST-AB 5 (2002) 074401
- [14] S. Heifets, G.V. Stupakov, S. Krinsky, PRST-AB 5 (2002) 064401

## 2.5 BEAM-BEAM EFFECTS

### 2.5.1 Beam-Beam Effects in Storage Ring Colliders

K. Hirata, Graduate U., Sokendai

We assume a two-ring  $e^+e^-$  collider with one interaction point (IP). Modifications for other types of colliders are straightforward. The beam-beam effects consist of (i) the beam-beam interaction at the IP ( $s = 0$ ) and (ii) the transformation from IP to IP. The two are independent of each other dynamically but all beam-beam phenomena are their combined effects [1, 2, 3, 4, 5].

#### 2.5.1.1 Infinitely short bunches

One can ignore the bunch length  $\sigma_z$  when  $\sigma_z \ll \beta_q$  at the IP ( $q$  stands for  $x$  or  $y$ ). Let  $\rho^\pm(x, y)$  be the transverse distribution of  $e^\pm$  beams. The luminosity (Sec.4.1) for very short bunches,  $\mathcal{L}_0$ , is

$$\mathcal{L}_0 = N^+N^-f_0 \int dx dy \rho^+(x, y)\rho^-(x, y) \quad (1)$$

At the IP, a particle is kicked by the beam-beam force

$$\begin{aligned} p_{x,y}^\pm &\rightarrow p_{x,y}^\pm - f_{x,y}^\pm(x, y) \\ f_{x,y}^\pm(x, y) &= -N_\mp \partial_{x,y} \phi^\pm(x, y; \rho^\mp) \end{aligned} \quad (2)$$

where  $\phi$  is the “potential” which depends on the  $\rho$  of the other beam,  $\phi^\pm(x, y; \rho) =$

$$\frac{r_e}{\gamma_\pm} \int dx^* dy^* G(x-x^*, y-y^*) \rho(x^*, y^*) \quad (3)$$

where  $G(x, y) = -\ln(x^2 + y^2)$ .

**Gaussian beams** If  $\rho^\pm$  is Gaussian,

$$\rho^\pm(x, y) = \frac{1}{2\pi\sigma_x^\pm\sigma_y^\pm} e^{-\frac{(x-\bar{x}^\pm)^2}{2(\sigma_x^\pm)^2} - \frac{(y-\bar{y}^\pm)^2}{2(\sigma_y^\pm)^2}}$$

where  $\bar{x}^\pm = \langle x \rangle_\pm$  and  $\sigma_x^\pm = \sqrt{\langle (x - \bar{x}^\pm)^2 \rangle_\pm}$ , etc, the force felt by a particle can be put in the Bassetti-Erskine formula [6],

$$f_{x,y}^\pm(x, y) = \frac{N_\mp r_e}{\gamma_\pm} F_{x,y}(x - \bar{x}^\mp, y - \bar{y}^\mp; \sigma_x^\mp, \sigma_y^\mp) \quad (4)$$

$$\begin{aligned} F_y(x, y, \sigma_x, \sigma_y) &+ iF_x(x, y, \sigma_x, \sigma_y) \\ &= \sqrt{\frac{2\pi}{\sigma_x^2 - \sigma_y^2}} \left[ w \left( \frac{x + iy}{\sqrt{2(\sigma_x^2 - \sigma_y^2)}} \right) \right. \\ &\quad \left. - \exp \left( -\frac{x^2}{2\sigma_x^2} - \frac{y^2}{2\sigma_y^2} \right) w \left( \frac{\frac{\sigma_y}{\sigma_x}x + i\frac{\sigma_x}{\sigma_y}y}{\sqrt{2(\sigma_x^2 - \sigma_y^2)}} \right) \right] \end{aligned} \quad (5)$$

Here  $w$  is the complex error function (Sec.1.5.1)

$$w(z) = e^{-z^2} \{1 - \text{erf}(-iz)\}$$

See also Sec.2.5.4.

Define the (incoherent) beam-beam parameter  $\xi^\pm$  as

$$\xi_q^\pm = \frac{N^\mp r_e}{\gamma_\pm} \frac{\beta_q^{0\pm}}{2\pi\sigma_q^\mp(\sigma_x^\mp + \sigma_y^\mp)} \quad (6)$$

where  $\beta_q^0$  is the *nominal* (= without beam-beam effect)  $\beta$ -function at the IP.

When both beams are Gaussian,

$$\mathcal{L}_0 = \mathcal{L}_0^G = \frac{N^+N^-f_0}{2\pi\Sigma_x\Sigma_y} e^{-\frac{(\bar{x}^+ - \bar{x}^-)^2}{2\Sigma_x^2} - \frac{(\bar{y}^+ - \bar{y}^-)^2}{2\Sigma_y^2}} \quad (7)$$

with the effective beam size

$$\Sigma_q = \sqrt{(\sigma_q^+)^2 + (\sigma_q^-)^2}$$

The coherent beam-beam parameter is

$$\Xi_q^\pm = \frac{N_\mp r_e}{\gamma_\pm} \frac{\beta_q^{0\pm}}{2\pi\Sigma_q(\Sigma_x + \Sigma_y)} \quad (8)$$

The luminosity, Eq.(7), can be expressed in terms of  $\Xi$  as

$$\begin{aligned} \mathcal{L}_0^G &= \frac{f_0}{r_e} \frac{N_+N_- \gamma_+ \gamma_-}{N_+ \gamma_+ + N_- \gamma_-} e^{-\frac{(\bar{x}^+ - \bar{x}^-)^2}{2\Sigma_x^2} - \frac{(\bar{y}^+ - \bar{y}^-)^2}{2\Sigma_y^2}} \\ &\times \left( \frac{\Xi_x^+}{\beta_x^{0+}} + \frac{\Xi_x^-}{\beta_x^{0-}} + \frac{\Xi_y^+}{\beta_y^{0+}} + \frac{\Xi_y^-}{\beta_y^{0-}} \right) \end{aligned} \quad (9)$$

which shows that the  $\beta^0$ 's should be small to obtain large luminosity, because there exists an empirical upper bound for  $\Xi$ 's.

**Linear dynamical effects** In the linear approximation, the dynamics can be treated as 1-D system. We denote  $(q, p_q)$  simply as  $(q, p)$  and omit  $\pm$  designations. Here we use the weak-strong picture, where one (weak) beam consists of a single particle so that it does not affect the other (strong) beam.

**tune shift** When  $\bar{x}^+ = \bar{x}^-$  and  $\bar{y}^+ = \bar{y}^-$  at the IP, the head-on collision reads

$$\begin{bmatrix} q \\ p \end{bmatrix} \rightarrow \mathbf{K} \begin{bmatrix} q \\ p \end{bmatrix}, \mathbf{K} = \begin{bmatrix} 1 & 0 \\ -\frac{4\pi\xi_0}{\beta_0} & 1 \end{bmatrix} \quad (10)$$

Here  $\xi_0$  and  $\beta_0$  are the nominal values. It is combined with the one-turn matrix without beam-beam effects,

$$\begin{bmatrix} q \\ p \end{bmatrix} \rightarrow \mathbf{M} \begin{bmatrix} q \\ p \end{bmatrix}, \mathbf{M} = \begin{bmatrix} \cos\mu_0 & \beta_0 \sin\mu_0 \\ -\frac{\sin\mu_0}{\beta_0} & \cos\mu_0 \end{bmatrix}$$

From the eigenvalues of  $\mathbf{MK}$ , we get

$$\cos \mu = \cos \mu_0 - 2\pi \xi_0 \sin \mu_0 \quad (11)$$

For small  $\xi_0$ , we have  $\nu = \mu/2\pi = \nu_0 + \xi_0$ . Thus  $\xi_0$  is also called the beam-beam tune shift parameter. Stability requires

$$\xi_0 \leq \frac{1}{2\pi} \cot \pi \nu_0 \quad (0 \leq \nu_0 < 1/2) \quad (12)$$

which is periodic in  $\nu_0$  with period 1/2.

Dynamic beta [1] The new  $\beta$ -function,  $\beta$ , at the IP can be found from

$$\mathbf{K}^{1/2} \mathbf{M} \mathbf{K}^{1/2} = \begin{bmatrix} \cos \mu & \beta \sin \mu \\ -\sin \mu / \beta & \cos \mu \end{bmatrix}$$

as

$$\beta = \beta_0 \frac{\sin \mu_0}{\sin \mu} = \frac{\beta_0}{\sqrt{1 + 4\pi \xi_0 \cot \mu_0 - 4\pi^2 \xi_0^2}} \quad (13)$$

Dynamic emittance [7] Synchrotron radiation effects can be included by putting

$$\begin{bmatrix} q \\ p \end{bmatrix} \rightarrow \Lambda \mathbf{M} \mathbf{K} \begin{bmatrix} q \\ p \end{bmatrix} + \begin{bmatrix} \hat{r}_1 \\ \hat{r}_2 \end{bmatrix} \quad (14)$$

where  $\Lambda$  is  $2 \times 2$  matrix representing the average effect of radiation on the one-turn matrix and the  $\hat{r}$ 's represent its stochastic effect. Define the envelope matrix

$$\mathbf{R} = \begin{bmatrix} \langle q^2 \rangle & \langle qp \rangle \\ \langle qp \rangle & \langle p^2 \rangle \end{bmatrix} \quad (15)$$

The equilibrium value of  $\mathbf{R}$  is the solution of

$$\mathbf{R} = (\Lambda \mathbf{M} \mathbf{K}) \mathbf{R} (\Lambda \mathbf{M} \mathbf{K})^t + \mathbf{D}, \quad \mathbf{D} = \langle \hat{r}_i \hat{r}_j \rangle \quad (16)$$

The equilibrium emittance is  $\epsilon = \sqrt{\det \mathbf{R}}$ .

$\Lambda$  and  $\mathbf{D}$  are determined by the lattice in the arc and do not depend on the beam-beam interaction. If

$$\Lambda = e^{-\delta} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \mathbf{D} = \epsilon_0 \sqrt{1-\lambda^2} \begin{bmatrix} \beta_0 & 0 \\ 0 & \beta_0^{-1} \end{bmatrix}$$

then

$$\epsilon = \frac{(1 + 2\pi \xi_0 \cot \mu_0) \epsilon_0}{\sqrt{1 + 4\pi \xi_0 \cot \mu_0 - 4\pi^2 \xi_0^2}} + O(\delta) \quad (17)$$

The dynamic beta, dynamic emittance and dynamic beam size ( $\sqrt{\epsilon \beta} + O(\delta)$ ) effects are illustrated in Fig.1. By Eqs.(12) and (13),  $\nu_0$  slightly above a half-integer is preferred.

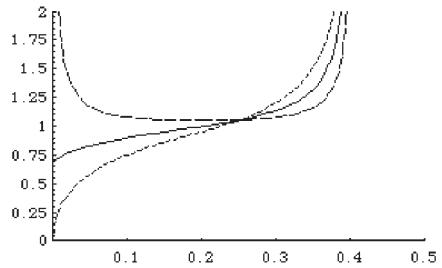


Figure 1: Dynamic beta, emittance and beam size effects as functions of  $\nu_0$  for  $\xi_0 = 0.05$ . Solid curve is  $\sigma/\sigma_0$ ; dashed curve is  $\epsilon/\epsilon_0$ ; dotted curve is  $\beta/\beta_0$ .

**Offset collision** When there are offsets ( $\bar{x}^+ - \bar{x}^- = d_x$  and  $\bar{y}^+ - \bar{y}^- = d_y$ ), the beam-beam kick, Eq.(10), is modified as [8]

$$\begin{bmatrix} x^\pm \\ p_x^\pm \\ y^\pm \\ p_y^\pm \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 0 & 0 & 0 \\ -k_{xx} & 1 & -k_{xy} & 0 \\ 0 & 0 & 1 & 0 \\ -k_{xy} & 0 & -k_{yy} & 1 \end{bmatrix} \begin{bmatrix} x^\pm \\ p_x^\pm \\ y^\pm \\ p_y^\pm \end{bmatrix} \quad (18)$$

where

$$\begin{aligned} k_{yy} + ik_{xy} &= \frac{\xi_0^\pm}{\beta_{0x}^\pm} \frac{4\pi\sigma_y}{\sigma_x - \sigma_y} \left[ 1 - e^{-A} \frac{\sigma_x}{\sigma_y} \right] \\ &\quad - \frac{\sqrt{\pi}}{\sqrt{2(\sigma_x^2 - \sigma_y^2)}} (d_y - id_x) \{ w(B) - e^{-A} w(C) \} \\ -ik_{xy} + k_{xx} &= \frac{\xi_0^\pm}{\beta_{0x}^\pm} \frac{4\pi\sigma_x}{\sigma_x - \sigma_y} \left[ 1 - e^{-A} \frac{\sigma_y}{\sigma_x} \right] \\ &\quad + \frac{\sqrt{\pi}}{\sqrt{2(\sigma_x^2 - \sigma_y^2)}} (-d_y + id_x) \{ w(B) - e^{-A} w(C) \} \end{aligned}$$

with  $A = \frac{d_x^2}{2\sigma_x^2} + \frac{d_y^2}{2\sigma_y^2}$ ,  $B = \frac{d_x + id_y}{\sqrt{2(\sigma_x^2 - \sigma_y^2)}}$  and  $C = \frac{\sigma_y d_x + i \sigma_x d_y}{\sqrt{2(\sigma_x^2 - \sigma_y^2)}}$ .

### Nonlinear 1-D models

Round beam For  $\sigma_x = \sigma_y$  (round beam) and  $\bar{x} = \bar{y} = 0$ , and if we observe a particle always with  $y = 0$ , we get

$$p_x^\pm \rightarrow p_x^\pm - f_r^\pm(x; \sigma_x^\mp) \quad (19)$$

$$f_r^\pm(x; \sigma_x) = \frac{2N_\mp r_e}{\gamma_\pm} \frac{1}{x} \left[ \exp \left( -\frac{x^2}{2\sigma_x^2} \right) - 1 \right]$$

Flat beam For  $\sigma_y \ll \sigma_x$  and  $y \ll \sigma_x$ , and we observe a particle always with  $x = 0$ , then [6]

$$p_y^\pm \rightarrow p_y^\pm - f_f^\pm(y; \sigma_y^\mp)$$

$$f_f^\pm(y; \sigma_y) = \frac{N_\mp r_e \sqrt{2\pi}}{\gamma_\pm} \frac{\sqrt{2\pi}}{\sigma_x^\mp} \operatorname{erf} \left( \frac{y}{\sqrt{2}\sigma_y} \right) \quad (20)$$

Both models show flip-flop hysteresis and cusp catastrophe behaviors [6].

**Nonlinear Perturbative Effects** Under the assumption of linear betatron oscillation and the beam-beam kick at the IP, the first order perturbation shows the amplitude dependence of the tune shifts [1]

$$\nu_x = \nu_x^0 + \xi_x^0 \left( \frac{1+1/a}{2} \right) \quad (21)$$

$$\times \int_0^\infty du \frac{Z_1\left(\frac{\alpha_x}{1+u}\right) Z_2\left(\frac{\alpha_y}{1+u/a^2}\right)}{(1+u)^{3/2}(1+u/a^2)^{1/2}}$$

$$\nu_y = \nu_y^0 + \xi_y^0 \left( \frac{1+a}{2} \right) \quad (22)$$

$$\times \int_0^\infty du \frac{Z_1\left(\frac{\alpha_y}{1+u}\right) Z_2\left(\frac{\alpha_x}{1+a^2u}\right)}{(1+u)^{3/2}(1+a^2u)^{1/2}}$$

Here

$$Z_1(x) = e^{-x}[I_0(x) - I_1(x)]$$

$$Z_2(x) = e^{-x}I_0(x)$$

where  $\sqrt{\alpha_q}$  is the amplitude normalized by  $2\sigma_q$ ,  $a = \sigma_y/\sigma_x$  is the aspect ratio, and  $I_0$  and  $I_1$  are modified Bessel functions.

**Coherent beam-beam motion** Assume the Rigid Gaussian Model (RGM) in which the beams are Gaussian with the nominal beam sizes and only their barycenter can change. The beam-beam kick for the barycenters are [6]

$$\bar{p}_q^\pm \rightarrow \bar{p}_q^\pm - \frac{N^\mp r_e}{\gamma_\pm} F_q(\bar{x}^\pm - \bar{x}^\mp, \bar{y}^\pm - \bar{y}^\mp; \Sigma_x, \Sigma_y) \quad (23)$$

where  $F$  is in Eq.(5). A consequence is that the linear approximation of the kick for  $\bar{p}_q$  is

$$\delta \bar{p}_q^\pm = -4\pi \Xi_0^\pm / \beta_0^\pm (\bar{q}^\pm - \bar{q}^\mp)$$

instead of Eq.(10).

The one-turn map is

$$\mathbf{M} = \hat{\mathbf{U}}(\nu_0^+, \nu_0^-) \mathbf{R}(\Xi_0^+, \Xi_0^-) \quad (24)$$

$\hat{\mathbf{U}}$  is the block-diagonal  $4 \times 4$  matrix through the arc,

$$\begin{aligned} \hat{\mathbf{U}}(\nu_0^+, \nu_0^-) &= \begin{bmatrix} \mathbf{U}(\nu_0^+) & \mathbf{0} \\ \mathbf{0} & \mathbf{U}(\nu_0^-) \end{bmatrix} \\ \mathbf{U}(\nu) &= \begin{bmatrix} \cos \mu & \sin \mu \\ -\sin \mu & \cos \mu \end{bmatrix} \end{aligned}$$

$\mathbf{M}$  operates on the dynamical variables  $(\mathbf{Z}_+, \mathbf{Z}_-)^t$  where

$$\mathbf{Z}_\pm = \sqrt{N_\pm \gamma_\pm} \begin{bmatrix} \bar{q}^\pm / \sqrt{\beta_0^\pm} \\ \sqrt{\beta_0^\pm} \bar{p}_q^\pm \end{bmatrix}$$

The beam-beam kick matrix  $\mathbf{R}$  is

$$\begin{aligned} \mathbf{R}(\Xi_z^+, \Xi_z^-) &= \begin{bmatrix} \mathbf{I} - \mathbf{A}(\Xi_z^+) & \mathbf{A}(\sqrt{\Xi_z^+ \Xi_z^-}) \\ \mathbf{A}(\sqrt{\Xi_z^+ \Xi_z^-}) & \mathbf{I} - \mathbf{A}(\Xi_z^-) \end{bmatrix} \\ \mathbf{A}(\Xi) &= \begin{bmatrix} 0 & 0 \\ 4\pi \Xi & 0 \end{bmatrix} \end{aligned} \quad (25)$$

The perturbed tunes are given by

$$\cos \mu = \frac{1}{2}(\cos \mu_0^+ + \cos \mu_0^-) \quad (26)$$

$$- \pi \Xi_0^+ \sin \mu_0^+ - \pi \Xi_0^- \sin \mu_0^- \pm \frac{1}{2} \sqrt{D}$$

where

$$D = 16\pi^2 \Xi_0^- \Xi_0^+ \sin \mu_0^+ \sin \mu_0^-$$

$$+ [\cos \mu_0^+ - \cos \mu_0^- - 2\pi \Xi_0^+ \sin \mu_0^+ + 2\pi \Xi_0^- \sin \mu_0^-]^2$$

The mapping is unstable if (i)  $\nu_0^+$  or  $\nu_0^-$   $\lesssim$  half-integers (half integer resonances); or (ii)  $\nu_0^+ + \nu_0^- \lesssim$  integer (sum resonances) [6].

When  $\nu_0^+ = \nu_0^- \equiv \nu_0$ ,  $\mathbf{M}$  is simply

$$\mathbf{M} = \begin{bmatrix} \mathbf{U} - \mathbf{U}\mathbf{A} & \mathbf{U}\mathbf{A} \\ \mathbf{U}\mathbf{A} & \mathbf{U} - \mathbf{U}\mathbf{A} \end{bmatrix}$$

which can be diagonalized as

$$\mathbf{LML}^{-1} = \begin{bmatrix} \mathbf{U} & \mathbf{O} \\ \mathbf{O} & \mathbf{U} - 2\mathbf{U}\mathbf{A} \end{bmatrix}$$

with

$$\mathbf{L} = \frac{1}{\sqrt{2}} \begin{bmatrix} \mathbf{I} & \mathbf{I} \\ -\mathbf{I} & \mathbf{I} \end{bmatrix}$$

Thus the dynamics is split into two modes:

(i)  $\sigma$ -mode: the mode is

$$\mathbf{Z}_\sigma = \frac{1}{\sqrt{2}}(\mathbf{Z}^+ + \mathbf{Z}^-)$$

whose eigentune is unaffected ( $\nu_\sigma = \nu_0$ ).

(ii)  $\pi$ -mode: the mode is

$$\mathbf{Z}_\pi = -\frac{1}{\sqrt{2}}(\mathbf{Z}^+ - \mathbf{Z}^-)$$

whose eigentune is gives by

$$\cos \mu_\pi = \cos \mu_0 - 2\pi(\Xi_0^+ + \Xi_0^-) \sin \mu_0$$

When two beams are identical,  $\Xi = \xi/2$  so that  $\nu_\pi \simeq \nu_0 + \xi_0$ .

Choice of working point In more general cases where there are  $N e^+$  bunches and  $N e^-$  bunches, periodically placed and colliding at  $N_{IP}$  periodically placed collision points (denoted as  $N \oplus N = N_{IP}$ ), we know that [10]: (i) All the cases can be reduced to one of the two irreducible cases,  $N_{IP} = 2N$  and  $N_{IP} = N$  with  $N$  odd; (ii)

In irreducible cases, with perfect symmetry between bunches, beams, IPs, and arcs between IPs, the instability occurs for  $\nu_0 \lesssim \ell N_{IP}/(2N)$  with  $\ell = 1, 2, \dots, 2N$  (This pattern repeats for  $\nu_0$  at every  $N_{IP}$ ); (iii) With broken symmetries between bunches and IPs, the instability pattern for  $\nu_0$  is similar to that of the  $1 \oplus 1 = 1$  case. Examples for  $3 \oplus 3=3$  and  $3 \oplus 3=6$  cases are shown in Fig.2.

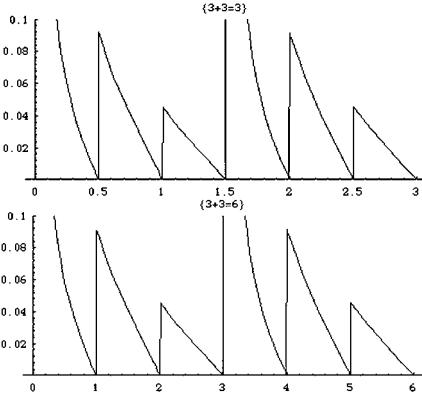


Figure 2: The threshold values of  $\Xi$  as functions of  $\nu_0$  (the “one turn” tune). Upper:  $3 \oplus 3=3$  case. Lower:  $3 \oplus 3=6$  case.

Spontaneous beam separation, etc. The RGM assumes each beam behaves as a single particle. Other features related to RGM are

- (i) an explanation why the four-beam compensation scheme of DCI experiment failed [11].
- (ii) showing serious instability if the two collider rings have different circumferences [6].
- (iii) beyond the threshold of the half integer resonances, two beams tend to separate from each other at the IP (Spontaneous Separation of Beams) [6].
- (iv) analysis of parasitic collisions [12, 13].
- (v) extension of Eq.(23) for the case where the transverse axes of both beams are tilted [14].

Visible frequencies When we deflect beams with a single frequency, we can observe peaks in its response corresponding to the above two modes [6]. When  $\Xi_0^\pm \ll 1$  the peaks sit on  $\nu_0$  and  $\nu_0 + \Xi_0^+ + \Xi_0^- \simeq \nu_0 + \xi_0$ . This is a result of the RGM. A more accurate estimate of the visible tune shift can be summarized by [6]

$$\nu_\pi = \nu_0 + G(r)\xi_0 \quad (27)$$

where  $r = \sigma_y/(\sigma_x + \sigma_y)$  and

$$G = \begin{cases} \Lambda(r), & \text{horizontal} \\ \Lambda(1-r), & \text{vertical} \end{cases} \quad (28)$$

where

$$\Lambda(r) = 1.33 - 0.37r + 0.279r^2 \quad (29)$$

For a very flat beam,  $r = 0$  and  $G$  is 1.33 (horizontal) and 1.24 (vertical).

For higher order coherent motions, refer to [2, 6] and references quoted therein. When the deflection is large, one can observe a hysteresis in the frequency response of the beam [6].

### 2.5.1.2 Long bunches

When  $\sigma_z$  is comparable to  $\beta_q$ , we should consider the bunch length effect (hour-glass effect) [15]. The interaction now depends also on the longitudinal position  $z$ . The collision between particles with  $z^+$  and  $z^-$  takes place at

$$s = S(z^+, z^-) = (z^+ - z^-)/2 \quad (30)$$

where the beam sizes are different from those at  $s = 0$  (the IP).

Luminosity The geometric reduction of luminosity due to the hour glass effect is discussed in Sec.4.1.

Dynamics The only known symplectic beam-beam map usable for a long bunch is the synchro-beam mapping [16]. The map for the collision of an  $e^+$  having  $z^+$  and a 3-D Gaussian  $e^-$  bunch is as follows: The strong  $e^-$  beam is cut into longitudinal slices: each slice is represented by  $z^-$ , its longitudinal coordinate at  $s = 0$ . Let  $n_i^-$  be the number of electrons in the  $i$ -th slice. Then the beam-beam kick is represented by the Lie operator

$$\prod_i \exp [n_i^- \phi(X, Y, \sigma_x^-(S), \sigma_y^-(S))] \quad (31)$$

which acts on the 6 phase space coordinates of the  $e^+$  at  $s = 0$ . Here,  $S$  is defined by (30) and

$$X = x + p_x S(z^+, z^-), \quad Y = y + p_y S(z^+, z^-)$$

The  $\sigma$ 's should be evaluated at  $s = S$  so that, e.g.,

$$\sigma_x^-(S) = \sqrt{\langle x^2 \rangle_0 + 2\langle xp_x \rangle_0 S + \langle p_x^2 \rangle_0 S^2}$$

Here  $\langle \cdot \rangle_0$  is the average at  $s = 0$ .

The map Eq.(31) is the successive application of the following map:

$$\begin{aligned} x^{\text{new}} &= x + S(z, z_*) f_X \\ p_x^{\text{new}} &= p_x - f_X \\ y^{\text{new}} &= y + S(z, z_*) f_Y \\ p_y^{\text{new}} &= p_y - f_Y \\ z^{\text{new}} &= z \\ p_z^{\text{new}} &= p_z - \frac{1}{2} f_X [p_x - \frac{1}{2} f_X] \\ &\quad - \frac{1}{2} f_Y [p_y - \frac{1}{2} f_Y] - g \end{aligned} \quad (32)$$

where

$$\begin{aligned}
 f_X &= \frac{n_i^- r_e}{\gamma_0} f_x(X, Y; \sigma_x(S), \sigma_y(S)) \\
 f_Y &= \frac{n_i^- r_e}{\gamma_0} f_y(X, Y; \sigma_x(S), \sigma_y(S)) \\
 g &= \frac{n_i^- r_e}{\gamma_0} [R_{22}(0; z^*) g_x(X, Y, \sigma_x(S), \sigma_y(S)) \\
 &\quad + R_{44}(0; z^*) g_y(X, Y, \sigma_x(S), \sigma_y(S))] S \\
 g_x(x, y, \sigma_x, \sigma_y) &= -\frac{1}{2(\sigma_x^2 - \sigma_y^2)} \\
 &\times \left\{ x f_x + y f_y + 2 \left[ \frac{\sigma_y}{\sigma_x} e^{-\left(\frac{x^2}{2\sigma_x^2} + \frac{y^2}{2\sigma_y^2}\right)} - 1 \right] \right\} \\
 g_y(x, y, \sigma_x, \sigma_y) &= \frac{1}{2(\sigma_x^2 - \sigma_y^2)} \\
 &\times \left\{ x f_x + y f_y + 2 \left[ \frac{\sigma_x}{\sigma_y} e^{-\left(\frac{x^2}{2\sigma_x^2} + \frac{y^2}{2\sigma_y^2}\right)} - 1 \right] \right\} \tag{33}
 \end{aligned}$$

### 2.5.1.3 Dispersion at IP, crossing angle

These two are known to introduce synchrobetatron resonances (Sec.2.3.4).

**Dispersion** Here, we review the properties of the arc, which are represented by the one-turn map around the IP ( $s = 0$ ). The revolution matrix  $\mathbf{M}(0)$  can be obtained by any 3-D tracking code but usually it is useful to represent it in terms of several parameters (3, 10, 21 for 1-D, 2-D, 3-D). One way to do it is as follows [17]:

$$\begin{aligned}
 \mathbf{M}(0) &= \mathbf{H}^{-1} \mathbf{R}^{-1} \begin{bmatrix} \mathbf{m}_u & 0 & 0 \\ 0 & \mathbf{m}_v & 0 \\ 0 & 0 & \mathbf{m}_w \end{bmatrix} \mathbf{R} \mathbf{H} \\
 \mathbf{H} &= \begin{bmatrix} \left\{ 1 - \frac{\det H_x}{1+a} \right\} \mathbf{I} & \frac{1}{1+a} \mathbf{H}_x \tilde{\mathbf{H}}_y & -\mathbf{H}_x \\ \frac{1}{1+a} \mathbf{H}_y \tilde{\mathbf{H}}_x & \left\{ 1 - \frac{\det H_y}{1+a} \right\} \mathbf{I} & -\mathbf{H}_y \\ -\tilde{\mathbf{H}}_x & -\tilde{\mathbf{H}}_y & a\mathbf{I} \end{bmatrix} \tag{34}
 \end{aligned}$$

Here

$$\begin{aligned}
 \mathbf{H}_q &= \begin{bmatrix} \zeta_q & \eta_q \\ \zeta'_q & \eta'_q \end{bmatrix}, \quad \tilde{\mathbf{H}}_q = \begin{bmatrix} -\eta'_q & \eta_q \\ \zeta'_q & -\zeta_q \end{bmatrix} \\
 a &= \sqrt{1 - \det H_x - \det H_y} \\
 \mathbf{R} &= \begin{bmatrix} b\mathbf{I} & -\tilde{\mathbf{R}}_2 & 0 \\ \tilde{\mathbf{R}}_2 & b\mathbf{I} & 0 \\ 0 & 0 & \mathbf{I} \end{bmatrix}, \quad b = \sqrt{1 - \det R_2} \\
 \mathbf{m}_l &= \begin{bmatrix} \cos \mu_l + \alpha_l \sin \mu_l & \beta_l \sin \mu_l \\ -\gamma_l \sin \mu_l & \cos \mu_l - \alpha_l \sin \mu_l \end{bmatrix}
 \end{aligned}$$

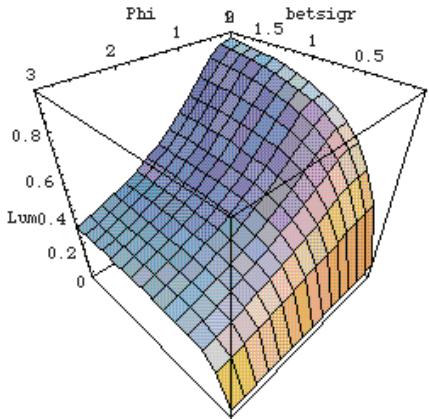


Figure 3: The geometrical luminosity reduction factor  $\mathcal{L}/\mathcal{L}_0$  as a function of  $\Phi \equiv \phi \sigma_z / \sigma_x$  and  $\beta_y / \sigma_z$ .

Here  $l$  stands for  $u, v, w$  and  $\gamma_l = (1 + \alpha_l^2)/\beta_l$ .  $\mathbf{R}_2$  is a  $2 \times 2$  horizontal-vertical coupling matrix defined and calculated from the physical one-turn matrix [17]. This factorization can express any symplectic matrix. The  $\eta$ 's (generalized energy dispersion) reduce to the usual dispersion when there is no dispersion in cavities; otherwise, the  $\zeta$ 's (generalized time dispersion) may appear [18].

The simplest case where there is a vertical dispersion at the IP and nothing else, i.e., where  $\eta'_y = \eta_x = \eta'_x = 0$ ,  $\alpha$ 's vanish,  $\zeta$ 's vanish, the beam-beam force is linear as  $\mathbf{K}$ , Eq.(10), and in the weak-strong situation, then the synchrotron tune is modified as [19]

$$\nu_z^0 \longrightarrow \nu_z^0 + \xi_y^0 \chi, \quad \chi = \frac{(\eta^0)^2 \sigma_{p_z}}{\sigma_z \beta_y^0} \tag{35}$$

as well as the usual transverse tune shifts.

**Crossing angle** The crossing angle can be treated by using the Lorentz boost for both beams. At the IP ( $s = 0$ ), we perform

$$\vec{x}(0) \xrightarrow{\mathcal{D}} \vec{x}^*(0^*) \xrightarrow{\text{BB}} \vec{x}^{**}(0^*) \xrightarrow{\mathcal{D}^{-1}} \vec{x}'(0) \xrightarrow{\mathcal{A}} \vec{x}(0) \dots$$

Here “BB” is the (symplectic) beam-beam kick for the head-on collision. In the ultrarelativistic limit,  $\mathcal{D}$  can be written as [20]

$$\begin{aligned}
 x^* &= \tan \phi z + (1 + h_x^* \sin \phi) x \\
 y^* &= y + h_y^* x \sin \phi \\
 z^* &= z/\cos \phi + h_z^* x \sin \phi \\
 p_x^* &= (p_x - h \tan \phi)/\cos \phi \\
 p_y^* &= p_y / \cos \phi \\
 p_z^* &= p_z - p_x \tan \phi + h \tan^2 \phi
 \end{aligned}$$

## Sec.2.5: BEAM-BEAM EFFECTS

Here  $\phi$  is the half crossing angle and the horizontal crossing was assumed,

$$h(p_x, p_y, p_z) = p_z + 1 - \sqrt{(p_z + 1)^2 - p_x^2 - p_y^2} \quad (36)$$

and  $h_q^* = \partial h^*/\partial p_q^*$ . See also Eqs.(15-19), Sec.2.3.4.

The geometrical luminosity reduction factor can be obtained from Eq.(6), Sec.4.1. This factor is shown in Fig.3 for the case with two identical beams.

## References

- [1] A.W. Chao, AIP Proc.127 (1983) p.201
- [2] A.W. Chao, SSCL-346 (1991)
- [3] AIP Proc. 214 (1990)
- [4] Proc. 3rd ICFA Beam Dynamics Workshop on Beam-Beam Effects in Circular Colliders (1989)
- [5] Proc. 7th ICFA Beam Dynamics Workshop on Beam-Beam Issues for Multibunch, High Luminosity Circular Colliders, JINR (1996)
- [6] K. Hirata, [3] p.175
- [7] K. Hirata, F. Ruggiero, PA 28 (1990) 137
- [8] B.W. Montague, CERN/ISR-GS/75-36 (1975)
- [9] A. Chao, AIP Proc. 57 (1979) p.42
- [10] K. Hirata, E. Keil, PA 56 (1996) 13
- [11] E. Keil, Proc. 3rd Advanced ICFA Beam Dynamics Workshop (1989) p.85
- [12] K. Hirata, AIP Proc. 214 (1990) p.441
- [13] M. Furmann et al, LBL-32435, ESG-193, DAPNIA/SPP 92-03, ABC-49 (1992)
- [14] V. Ziermann, SLAC-PUB-5479 (1991)
- [15] S. Krishnagopal, R. Siemann, PR D41 (1990) 2312
- [16] K. Hirata, H. Moshammer, F. Ruggiero, PA 40 (1993) 205
- [17] K. Ohmi, K. Hirata, K. Oide, PR E49 (1994) 751
- [18] S. Petracca, K. Hirata, PAC 97
- [19] S. Petracca, K. Hirata, AIP Proc. 395 (1997) p.369
- [20] K. Hirata, PRL 74 (1995) 2228

### 2.5.2 Crab Waist Collision Scheme

M. Zobov, INFN LNF

The crab waist (CW) collision scheme was proposed [1, 2] and successfully tested at the  $\Phi$ -factory DAΦNE [3]. At present this scheme is considered to be most attractive for the next generation lepton factories since it holds the promise of increasing the luminosity of the storage-ring colliders by 1-2 orders of magnitude beyond the current state-of-art, without any significant increase in beam current and without reducing the

bunch length (see Sec.4.13 for different collisions schemes).

The CW scheme can substantially increase collider luminosity since it combines several potentially advantageous ideas. Consider two bunches colliding under a horizontal crossing angle  $\theta$  (Fig.1). The CW principle can be explained in three steps. The first step is large Piwinski angle  $\Phi = (\sigma_z/\sigma_x) \tan(\theta/2) \gg 1$ . In the CW scheme,  $\Phi$  is increased by decreasing  $\sigma_x$  and increasing  $\theta$ . In this way we can gain in luminosity and the horizontal tune shift decreases (as discussed in Sec.4.13); parasitic collisions (PC) (Sec.2.5.4) become negligible since with larger  $\theta$  and smaller  $\sigma_x$  the beam separation at the PC is larger in terms of  $\sigma_x$ . But the most important effect is that the overlap area of the colliding bunches is reduced, since it is proportional to  $\sigma_x/\theta$  (see Fig.1).

As the second step, the vertical  $\beta$ -function  $\beta_y^*$  can be made comparable to the overlap area size (i.e.  $\ll \sigma_z$ ),

$$\beta_y^* \approx \frac{2\sigma_x}{\theta} \cong \frac{\sigma_z}{\Phi} \ll \sigma_z$$

So, reducing  $\beta_y^*$  at the IP gives us several advantages:

- Luminosity increase with the same bunch current.

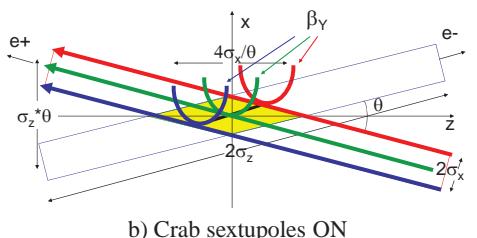
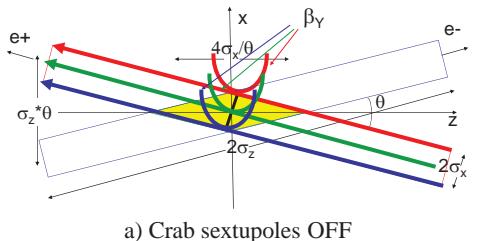


Figure 1: Crab Waist collision scheme.

- Possibility of the bunch current increase (if it is limited by  $\xi_y$ ), thus farther increasing the luminosity.
- Suppressing vertical synchrobetatron resonances [4].

Furthermore, there is no need in decreasing the bunch length to increase the luminosity as required in standard collision schemes (Sec.4.13). This will help solve the problems of HOM heating, coherent synchrotron radiation of short bunches and excessive power consumption.

However, implementation of these two steps introduces new beam-beam resonances which may strongly limit the maximum achievable tune shifts. At this point the crab waist transformation [1, 2] enters the game. This is the third step. As seen in Fig.1b, the  $\beta$ -function waist of one beam is oriented along the central trajectory of the other beam. In practice the CW vertical  $\beta$ -function rotation is provided by sextupoles placed on both sides of the IP in phase with the IP horizontally and at  $\pi/2$  phase vertically (Fig.2).

The integrated strength  $K$  of these sextupoles should satisfy a condition depending on  $\theta$  and the  $\beta$ -functions at the IP and the sextupole locations,

$$K = \frac{1}{\theta} \frac{1}{\beta_y^* \beta_y} \sqrt{\frac{\beta_x^*}{\beta_x}}$$

The crab waist transformation gives a small geometric luminosity gain due to the vertical  $\beta$ -function redistribution along the overlap area. It is estimated to be of the order of several percent. However, the dominating effect comes from the suppression of betatron and synchrobetatron resonances (Sec.2.3.4) arising (in collisions without CW) due to the vertical motion modulation by the horizontal betatron oscillations [5]. The collision of flat beams with  $\sigma_y \ll \sigma_x$  is the essential condition for the resonance suppression with the crab sextupoles [6].

Fig.3 demonstrates the resonance suppression applying the frequency map analysis (FMA) for the beam-beam interaction in DAΦNE in CW collisions [6]. The crab waist collision test at

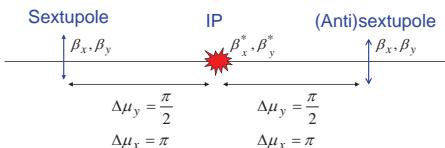


Figure 2: Crab sextupole locations.

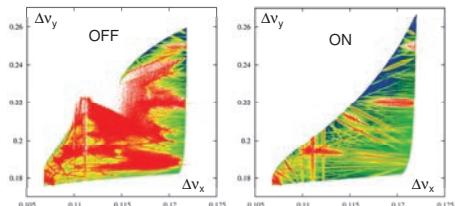


Figure 3: DAΦNE beam-beam footprint with crab sextupoles off (left) and on (right) obtained by FMA [6].

DAΦNE has provided a factor of 3 luminosity increase [3] in good agreement with numerical simulations [7].

All the projects of the next generation lepton factories are based on the crab waist scheme. Their designs rely on collisions of flat beams with very small emittances [8] comparable to those obtained in modern synchrotron light sources. So issues of achieving and preserving small emittances at high beam current become of primary importance. This requires a careful study of emittance growth due to wake fields, e-clouds, ion-related phenomena, intrabeam scattering etc.

## References

- [1] P. Raimondi, 2nd SuperB Workshop, 2006
- [2] P. Raimondi et al, e-Print:physics/0702033 (2007)
- [3] M. Zobov et al, PRL104, 174801 (2010)
- [4] D.V. Pestrikov, NIM A336, 427 (1993)
- [5] P. Raimondi, D. Shatilov, M. Zobov, EPAC08, p.2620
- [6] D. Shatilov et al, PRST-AB 14, 014001 (2011)
- [7] M. Zobov et al, ICFA Beam Dyn. Newslett. 48, p.34
- [8] ICFA Beam Dynamics Letter 48 (2009)

### 2.5.3 Beam-Beam Effects in Linear Colliders

*P. Chen, Taiwan U. & SLAC  
D. Schulte, CERN*

In linear colliders, the colliding beams must have small transverse dimensions to reach high luminosity. This leads to a strong electromagnetic force exerted by each beam on the particles of the other. This gives rise to strong disruption effects and leads to the emission of beamstrahlung. A full analytic treatment of the beam-beam interaction is usually not possible, so it is most often studied using simulation codes, e.g. CAIN [1] and GUINEA-PIG [2]; for the latter also a C++ version exists [3].

## Sec.2.5: BEAM-BEAM EFFECTS

The beams in linear colliders are normally flat,  $\sigma_x \gg \sigma_y$ . We imply this in the following; the reason for this choice is beamstrahlung.

### 2.5.3.1 Disruption

**Crossing angle** Often a crossing angle  $\theta_c$  is foreseen in the horizontal plane between the two colliding beam lines. The resulting luminosity loss is avoided by using a crab crossing scheme (Sec.4.1). In this scheme an  $x'-z$  correlation is introduced in the beam before the collision, such that it is aligned with the longitudinal axis of the laboratory system at the collision, see Fig.1. For this case the beam-beam effects can be treated as if no crossing angle were present.

**Disruption parameter** We first consider the motion of a single particle moving parallel to the beam axis near  $x = y = 0$  in the field of the oncoming bunch assumed to have a Gaussian profile with rms sizes  $\sigma_{x,y,z}$ .

Important are the *disruption parameters*

$$\mathcal{D}_{x,y} = \frac{2Nr_e\sigma_z}{\gamma\sigma_{x,y}(\sigma_x + \sigma_y)} \quad (1)$$

where  $N$  is the bunch particle number,  $r_e$  the classical electron radius,  $\gamma$  the Lorentz energy factor. For  $\mathcal{D}_{x,y} \ll 1$  particles close to the beam axis will be deflected toward the axis

$$x' = \mathcal{D}_x \frac{x}{\sigma_z}, \quad y' = \mathcal{D}_y \frac{y}{\sigma_z}$$

If  $\mathcal{D} \gg 1$  the particle will move transversely during the passage through the oncoming bunch and the description of the beam-beam interaction becomes complex.

**Luminosity enhancement** Since the particles of each beam are focused by the other beam, the beam sizes decrease during collision. This increases luminosity by the luminosity enhancement factor  $H_D$ , which is typically in the range 1–2. While formulae exist that approximate  $H_D$  as a function of beam parameters, it is generally calculated using one of the simulation codes.

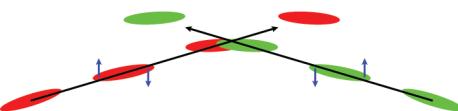


Figure 1: Conceptual sketch of the crab crossing scheme.

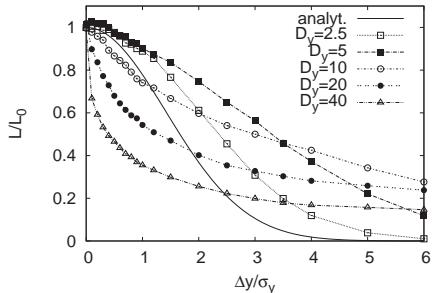


Figure 2: Luminosity as a function of beam-beam offset for different disruption parameters. The simple analytic expression is compared to simulation with GUINEA-PIG.

**Kink instability** The dependence of the luminosity on the beam-beam offset is strongly affected by the beam-beam interaction. For rigid bunches and assuming  $\beta_{x,y} \gg \sigma_z$ , the luminosity as a function of beam-beam jitter  $\Delta_{x,y}$  has

$$\mathcal{L} = \mathcal{L}_0 \exp \left( -\frac{\Delta_x^2}{4\sigma_x^2} - \frac{\Delta_y^2}{4\sigma_y^2} \right) \quad (2)$$

For limited disruption parameters, the luminosity is less sensitive to offsets since the beams attract each other. But for larger vertical disruption parameters the collisions become more complex, see Fig.2. This is due to the development of a two-stream instability during the collision. Such an instability can even occur if the mean offset of the beam particles is zero but if the different longitudinal slices have offsets. In this case the projected beam sizes are not a good measure for the achieved luminosity. By adjusting the mean offset and angle of the incoming beam slightly the luminosity can however be recovered [4].

**Multi-bunch kink instability** Parasitic crossing of the outgoing and incoming bunches can lead to a multi-bunch kink instability. If bunches collide with a small vertical offset, they receive a strong vertical deflection. After the collision these bunches can have parasitic crossings with the next ones of the incoming beams and because of their now large vertical offset can kick them vertically. In the likely case that all bunches in the trains have the same offset  $\Delta y_0$  this effect will increase the offset at the collision point along the train. In case of an infinitely long train and small offset this will lead to a maximum offset  $\Delta y$ ,

$$\Delta y = \frac{\Delta y_0}{1 - n_c \frac{4Nr_e}{\gamma\theta_c^2} \frac{\partial y'}{\partial \Delta y}} \quad (3)$$

Here,  $n_c$  is the number of parasitic crossings before the beams are separated,  $\theta_c$  is the full crossing angle and  $\partial y'/\partial \Delta y$  is the deflection at the IP. To limit the offset amplification to a factor two one requires

$$n_c \frac{4Nr_e}{\gamma \theta_c^2} \frac{\partial y'}{\partial \Delta y} < \frac{1}{2} \quad (4)$$

The slope  $\partial y'/\partial \Delta y$  of beam-beam deflection as a function of offset can be obtained from simulation. In order to obtain a first order of magnitude expression, one can simplify a formula in [19], which is valid for rigid beams (i.e. small disruption parameters), small vertical and no horizontal offset:

$$\frac{\partial y'}{\partial \Delta y} \approx \frac{1}{2} \frac{D_y}{\sigma_z} \quad (5)$$

This leads to the requirement:

$$n_c D_x D_y \frac{\sigma_x^2}{\sigma_z^2} \frac{1}{\theta_c^2} \approx n_c \frac{4N^2 r_e^2}{\gamma^2 \sigma_x \sigma_y} \frac{1}{\theta_c^2} < \frac{1}{2} \quad (6)$$

In case of high disruption the vertical angles tend to be smaller than indicated, so the effect will be smaller. Note that significant production of coherent pairs can lead to a stronger parasitic deflection of the beams, since the coherent pairs are deflected more than the beam particles due to their low energy.

### 2.5.3.2 Beamstrahlung

The bending of the trajectories during beam-beam interaction leads to production of beamstrahlung, which is photon emission similar to synchrotron radiation. In typical linear collider cases the number of photons emitted per beam particle is  $O(1)$ , so one has to consider the quantum nature of the process. The emitting particle will collide further, contributing to the luminosity at a lower than nominal center-of-mass energy.

**Beamstrahlung parameter** All beamstrahlung effects can be described by the Lorentz-invariant, dimensionless *beamstrahlung parameter*  $\Upsilon$ , defined as

$$\Upsilon \equiv \frac{e\hbar}{m_e^3 c^4} (p_\mu F^{\mu\lambda} p^\nu F_{\lambda\nu})^{1/2} \quad (7)$$

where  $p_\mu$  is the four-momentum of the particle and  $F_{\mu\nu}$  is the mean energy-momentum stress tensor of the beam field. In terms of the Schwinger critical field,

$$B_c \equiv \frac{m^2 c^3}{e\hbar} \approx 4.4 \times 10^{13} \text{Gauss} \quad (8)$$

and the mean field strength  $\langle E + B \rangle$ ,  $\Upsilon$  can be expressed as

$$\Upsilon = \gamma \frac{\langle E + B \rangle}{B_c} \approx \frac{5}{6} \frac{r_e^2 \gamma N}{\alpha \sigma_z (\sigma_x + \sigma_y)} \quad (9)$$

$\Upsilon$  is a measure of the field strength in the electron's rest frame, i.e.  $\gamma \langle E + B \rangle$ , in units of  $B_c$ . As  $B_c$  signals the onset of nonlinear QED effects,  $\Upsilon \ll 1$  corresponds to the classical limit, while  $\Upsilon \gg 1$  the deep quantum regime, of the beamstrahlung effects.

The fact that the beam field is not uniform across the beam indicates that the mean field strength introduced above may not be sufficient in describing certain beamstrahlung phenomena, and a local definition is necessary. The maximum beam field corresponds to a maximum  $\Upsilon_{\max}$ , which is related to the mean value by

$$\Upsilon_{\max} \approx \frac{12}{5} \Upsilon \quad (10)$$

In principle, all beamstrahlung effects should be calculated locally, and integrated over the beam cross section. However, the global beamstrahlung parameter, Eq.(7), often suffices, at least to understand the fundamental scaling of effects.

### Number of photons and average energy loss

The photon emission is described by the Sokolov-Ternov spectrum

$$\frac{\partial \dot{w}}{\partial \omega} = \frac{\alpha}{\sqrt{3}\pi\gamma^2} \left[ \int_x^{\infty} K_{\frac{5}{3}}(x') dx' + \frac{\hbar\omega}{E} \frac{\hbar\omega}{E - \hbar\omega} K_{\frac{2}{3}}(x) \right]$$

Here,  $x = \frac{\omega}{\omega_c} \frac{E}{E - \hbar\omega}$  and  $K_{5/3}$  and  $K_{2/3}$  are modified Bessel functions.

The most important physical parameter in beamstrahlung is the average number of photons emitted per electron during the collision,

$$\eta_\gamma \approx 2.54 \left[ \frac{\alpha^2 \sigma_z}{r_e \gamma} \Upsilon \right] U_0(\Upsilon) \quad (11)$$

$$U_0(\Upsilon) \approx \frac{1}{[1 + \Upsilon^{2/3}]^{1/2}}$$

The average fractional energy loss due to beamstrahlung is

$$\delta_B \approx 1.24 \left[ \frac{\alpha^2 \sigma_z}{r_e \gamma} \Upsilon \right] \Upsilon U_1(\Upsilon) \quad (12)$$

$$U_1(\Upsilon) \approx \frac{1}{[1 + (1.5\Upsilon)^{2/3}]^2}$$

The average energy of all the emitted photons is

$$\delta_\gamma = \frac{4\sqrt{3}}{15} \Upsilon \frac{U_1(\Upsilon)}{U_0(\Upsilon)} \approx \begin{cases} 0.462\Upsilon & (\Upsilon \rightarrow 0) \\ 0.254 & (\Upsilon \rightarrow \infty) \end{cases} \quad (13)$$

Note that  $\delta_\gamma$  is finite (about 1/4 of the initial electron energy) when  $\Upsilon \rightarrow \infty$ .

**Luminosity vs beamstrahlung** Beamstrahlung is one factor that limits luminosity in linear colliders. The luminosity spectrum affects the quality of the physics event analysis in the detector, hence one limits  $n_\gamma$ . The luminosity spectrum is also degraded by initial state radiation, which is the emission of photons as a part of the physics process. One usually tries to limit the spectrum degradation due to beamstrahlung to the level of degradation due to initial state radiation. This typically leads to a required  $n_\gamma = \mathcal{O}(1)$ .

Flat beams allow to maximize the luminosity, which is  $\propto 1/(\sigma_x \sigma_y)$  and at the same time limit the  $n_\gamma$ , which  $\propto 1/(\sigma_x + \sigma_y)$  in the classical case. In the following, we consider the classical case  $U_0 \approx U_1 \approx 1$  and flat beams  $\sigma_x \gg \sigma_y$ . The total luminosity is

$$\mathcal{L} = H_D \frac{N^2}{4\pi\sigma_x\sigma_y} n_b f_r$$

with  $n_b$  the number of bunches per train and  $f_r$  the repetition rate for the trains. Neglecting  $H_D$  and using  $\sigma_x + \sigma_y \approx \sigma_x$ , we find

$$\mathcal{L} \propto \frac{n_\gamma}{\sigma_y} \eta P_{wall}$$

where  $P_{wall}$  is the total power consumption of the collider and  $\eta$  the efficiency of transferring this power into beam power. The average energy of the photons is given by

$$E_\gamma \propto n_\gamma \frac{1}{\sigma_z}$$

The most relevant property of the luminosity spectrum is usually the fraction of luminosity in the peak close to the nominal center-of-mass energy. This peak depends on  $n_\gamma$  and to some extent on  $E_\gamma$ . Usually the relevant luminosity spectrum is derived by simulation.

**Luminosity vs. beamstrahlung in the non-classical case** In this case the radiation is suppressed to some extent since the critical energy is above the beam energy. For  $\Upsilon \gg 1$ , one can approximate  $U_0(\Upsilon)$  with  $\Upsilon^{-1/3}$ . This leads to the expression

$$\mathcal{L} \approx \frac{0.024}{\alpha^2} n_\gamma^{3/2} \frac{\sqrt{\Upsilon}}{\sqrt{\sigma_z r_e} \sigma_y} \frac{1}{N} N n_b f_r$$

In this case a shorter bunch suppresses the beamstrahlung, which allows to reduce  $\sigma_x$  to have the same number of beamstrahlung photons. Consequently, the luminosity can be increased for shorter bunches.

**Vertical beam size and luminosity** Limitations for  $\sigma_y$  can be of different nature. One limitation is the hour glass effect. If  $\beta_y < \sigma_z$ , the beam is focused to a small size but only over a small length, which reduces luminosity. Hence one requires  $\beta_y \approx \sigma_z$ . Applying this limit the luminosity can be rewritten for the classical case as

$$\mathcal{L} \propto \frac{n_\gamma}{\sqrt{\epsilon_y \sigma_z}} \eta P_{wall} \propto \frac{\sqrt{n_\gamma E_\gamma}}{\sqrt{\epsilon_y}} \eta P_{wall}$$

In the case of short bunches  $\beta_y$  can be limited by the design of the final focus system. Or the beam-beam jitter resulting from dynamic effects may pose a lower limit on beam size.

**Depolarization** There are two major mechanisms that cause the spin depolarization of the  $e^-$  and  $e^+$  beams during beam-beam interaction. One is the classical precession under the beam field described by the Thomas-BMT equation (Sec.2.6.1). The other is the spin-flip, or the Sokolov-Ternov, effect (Sec.2.6.6) upon the emission of beamstrahlung photons.

The average final depolarization due to the Thomas-BMT effect is [17]

$$\langle \Delta P \rangle_{BMT} = \frac{3}{50\pi^2} \left( \frac{n_\gamma}{U_0(\Upsilon)} \right)^2 \quad (14)$$

while that due to Sokolov-Ternov effect is [17]

$$\langle \Delta P \rangle_{ST} \approx \frac{7}{12} \Upsilon^2 n_\gamma \quad (1 \lesssim \Upsilon \lesssim 100) \quad (15)$$

The net average depolarization is then the sum of the two contributions. It is also important to estimate the luminosity weighted depolarization,  $[\Delta P]$ , which is related to  $\langle \Delta P \rangle$  by a simple numerical factor,

$$[\Delta P] = 0.273 \langle \Delta P \rangle \quad (16)$$

### 2.5.3.3 Background and spent beam

Beam-beam effects are an important source for background in the detectors.

**Disruption angles** The beam-beam disruption angles of the particle are important to define the apertures of the post collision line. The angle is characterized by the nominal deflecting angle

$$\theta_0 \equiv \frac{2r_e N}{\gamma(\sigma_x + \sigma_y)} = \frac{\mathcal{D}_x \sigma_x}{\sigma_z} = \frac{\mathcal{D}_y \sigma_y}{\sigma_z} \quad (17)$$

For a flat beam one can estimate the maximum deflection in the horizontal plane. In the limit of large  $\beta$ -functions and small disruption parameter one finds

$$\theta_{x,max} \approx 0.77 \theta_0$$

which is reached by particles at  $x \approx 1.3\sigma_x$ . In the vertical direction the maximum field is reached at  $y \approx 0.0065\sigma_x$  with

$$\theta_{y,max} \approx 1.25\theta_0$$

For head-on collision the number of particles that experience the full deflection in the vertical plane can be negligible. But collisions with a beam-beam offset can lead to a number of particles having large angles. Further note that particles may be deflected to larger angles if they radiate beamstrahlung and that the initial beam divergence needs to also be taken into account. In addition secondary particles, in particular coherent and incoherent pairs, discussed below, can be important for the post collision line aperture requirements.

**Coherent Pairs** A photon propagating through a transverse EM field has a probability of turning into  $e^+e^-$  pairs. This process has been called *coherent pair creation* [20]. Such a process can be contributed by either the real, beamstrahlung, photons, or the virtual photons comoving with the high energy particle. For beamstrahlung coherent pair creation, the average number of  $e^+e^-$  pairs created per primary electron is

$$n_{e^+e^-}^b = \left[ \frac{\alpha\sigma_z}{\lambda_c\gamma} \Upsilon \right]^2 \Xi(\Upsilon)$$

$$\Xi(\Upsilon) = \begin{cases} (7/128) \exp(-16/3\Upsilon), & (\Upsilon \lesssim 1) \\ 0.295\Upsilon^{-2/3}(\ln \Upsilon - 2.488) & (\Upsilon \gg 1) \end{cases}$$

The energy spectrum of the coherent pairs can be approximated by (when  $\Upsilon \lesssim 10$ )

$$\frac{dn_{e^+e^-}^b}{dx} = 0.2 \left[ \frac{\alpha\sigma_z}{\lambda_c\gamma} \Upsilon \right]^2 \sqrt{\frac{1-x}{\Upsilon x}} \times \exp \left[ -\frac{2}{3\Upsilon} \left( \frac{1-x}{x} + \frac{4}{1-x} \right) \right] \quad (18)$$

When  $\Upsilon \gg 1$ , the contribution from virtual photons becomes significant. The virtual process is sometimes called *trident cascade* [20],

$$n_{e^+e^-}^v = \left[ \frac{\alpha\sigma_z}{\lambda_c\gamma} \Upsilon \right] \Omega(\Upsilon)$$

$$\Omega(\Upsilon) \approx 0.23\alpha \ln \Upsilon \quad (\Upsilon \gg 1)$$

In the multi-TeV regime, coherent pair production can lead to significant pair production. The number of pairs can be a significant fraction of the number of beam particles. Particles from coherent pairs have significantly lower energies than the beam particles. The electron of a pair can fly in the direction of the electron beam, in which case

it is focused in the same way by the positrons. Or it can fly in the direction of the positrons, in which case it will be deflected to larger angles by the electron beam. The detector has to be protected from these pairs by carefully choosing the apertures for the spent beam, which can impact the choice of crossing angle between the two beam lines, and by introducing masking systems. Both beam-beam codes allow the simulation of coherent pairs from beamstrahlung, GUINEA-PIG also includes the trident cascade process.

**Incoherent pairs** Two colliding photons can produce an  $e^+e^-$  pair. These incoherent pairs have an important impact on the detector design. Most of the pair particles are produced with low energies and at small angles, but some can have larger angles at production. The particles are deflected by the beam fields in the same fashion as the coherent pairs and can hit the vertex detector and produce background that prevents proper event reconstruction. This is avoided by the use of a high solenoid field in the detector which confines the particles and by use of a large enough radius of the innermost layer of the vertex detector, which results in a limited vertex resolution. Beam-beam codes allow to study the generation of incoherent pairs and their deflection by the beams.

**Hadronic events** The collision of two photons can also produce hadrons. The cross section for this is significantly lower than for incoherent pair production, but the produced particles can impact the reconstruction of the physics events. Different parametrizations of the cross sections for hadronic events exist. For given beam parameters, the rates can be calculated with GUINEA-PIG, which also produces the information necessary to generate them for further analysis.

## References

- [1] P. Chen et al, NIM A 355, 107 (1995)
- [2] D. Schulte, DESY-TESLA-97-08 (1996)
- [3] D. Schulte et al, PAC07-THPMN010
- [4] R. Brinkmann, O. Napoly, D. Schulte, DESY-TESLA-2001-16, CLIC-Note-505
- [5] D. Schulte, CLIC-Note-560
- [6] R. Hollebeek, NIM 184 (1981) 333
- [7] P. Chen, K. Yokoya, PRD 38 (1988) 987
- [8] R.J. Noble, NIM A256 (1987) 427
- [9] M. Bell, J.S. Bell, PA 24 (1988) 1; R. Blankenbecler, S.D. Drell, PRL 61 (1988) 2324; P. Chen, K. Yokoya, PRL 61 (1988) 1101; M. Jacob, T.T.

## Sec.2.5: BEAM-BEAM EFFECTS

- Wu, Nucl. Phys. B303 (1988) 389; V.N. Baier, V.M. Katkov, V.M. Strakhovenko, Nucl. Phys. B328 (1989) 387
- [10] P. Chen, Lecture Notes in Phys. 296, Springer-Verlag (1988)
- [11] K. Yokoya, P. Chen, Lecture Notes in Phys. 400, Springer-Verlag (1992)
- [12] P. Chen, Nucl. Phys. B 51A (1996) 179
- [13] K. Yokoya, KEK Report 85-9 (1985)
- [14] P. Chen, PAC 93, p.617
- [15] K. Yokoya, P. Chen, PAC 89, p.1438
- [16] P. Chen, PRD 46 (1992) 1186
- [17] K. Yokoya, P. Chen, AIP Proc. 187 (1988) p.938
- [18] I.F. Ginzburg et al, NIM 205 (1983) 47
- [19] O. Napoly, B. Zotter, CLIC-Note 289 (1995)
- [20] P. Chen, in Research Directions of the Decade, Snowmass, World Scientific (1990); P. Chen, V. Telnov, PRL 63 (1990) 1796

### 2.5.4 Parasitic Beam-Beam Effects and Separation Schemes

J.M. Jowett, CERN

#### 2.5.4.1 Separation schemes

**Encounter points** Counter-rotating  $(+,-)$  bunches of equal speed, starting at azimuths  $s^+, s^-$ , at a given initial time, in a collider of circumference  $C$ , encounter each other (and may collide) at two azimuths separated by  $C/2$ ,

$$(\hat{s}, \tilde{s}) = \left( \frac{s^+ + s^-}{2}, \frac{C + s^+ + s^-}{2} \right) \pmod{C} \quad (1)$$

All encounters between beams of arbitrary bunch structure (filling scheme)  $(s_1^+, \dots, s_{M^+}^+)$ ,  $(s_1^-, \dots, s_{M^-}^-)$  are given by a matrix of such pairs

$$\begin{bmatrix} (\hat{s}, \tilde{s})_{11} & \dots & (\hat{s}, \tilde{s})_{1j} & \dots & (\hat{s}, \tilde{s})_{1M^-} \\ \vdots & \ddots & \vdots & & \vdots \\ (\hat{s}, \tilde{s})_{i1} & \dots & (\hat{s}, \tilde{s})_{ij} & \dots & (\hat{s}, \tilde{s})_{iM^-} \\ \vdots & \ddots & \vdots & & \vdots \\ (\hat{s}, \tilde{s})_{M^++1} & \dots & (\hat{s}, \tilde{s})_{M^++j} & \dots & (\hat{s}, \tilde{s})_{M^++M^-} \end{bmatrix} \quad (2)$$

The  $2M^-$  encounters experienced by bunch  $i$  of the “+” beam are listed in row  $i$  of the matrix. The  $s_i^\pm$  differ amongst themselves by multiples of the rf wavelength  $\lambda_{\text{rf}}$  and the  $\hat{s}_{ij}, \tilde{s}_{ij}$  by  $\lambda_{\text{rf}}/2$ . Unequal speeds or circumference can lead to moving encounter points.

Unwanted collisions can be suppressed entirely, by separating the beams into two separate vacuum chambers at some distance from the interaction point (IP), and/or partly, by separating the

orbits within a single chamber, over some part of the circumference. (Two rings with different circumferences and a common section are also possible.) Separation at the IP may be required, e.g., during injection at lower energy.

**Typical separation schemes** If the beam momentum-to-charge ratios are equal and opposite at the IP,  $\vec{p}^+/q^+ = \vec{p}^-/q^-$  (e.g., equal-energy  $e^+e^-$ -collider), orbit separation requires transverse electrostatic or time-dependent magnetic fields. Otherwise, static magnetic fields can also separate (e.g., unequal-energy  $e^+e^-$ -collider, pp-collider). If there are no common sections of the orbits (e.g., two-ring collider with crossing angle) a separating field may not be necessary.

Near an IP, design of a separation scheme is intimately connected with interaction region design [1, 2, 3, 4, 5, 6]; see Figs.1-2 for some typical schemes. Schemes with crossing angles can allow smaller bunch spacings than those with head-on collisions and are an essential feature of most modern high-luminosity colliders.

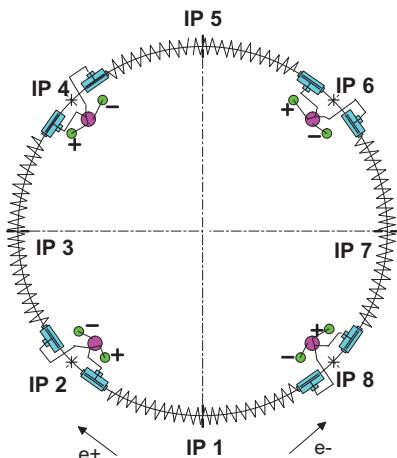


Figure 1: Horizontal orbit of one beam in a “pretzel” scheme in equal-energy  $e^+e^-$ -collider LEP with many potential locations for parasitic encounters in the arcs where the beams are separated in a single chamber. Head-on collisions occur at the even-numbered IPs. Anti-symmetry about the IPs helps to equalize the global optical parameters between the two beams. (Courtesy J.-P. Deluen)

**Side-effects** Separation schemes which involve displacements of the ideal orbit from a nominal orbit (typically passing through the centers of

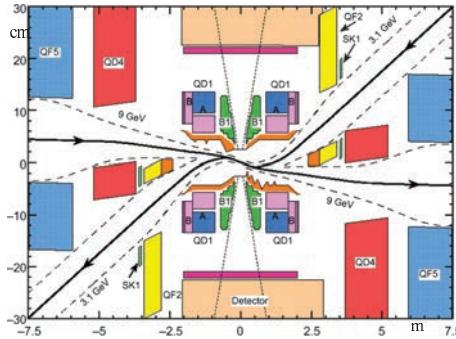


Figure 2: Differential bending separation by magnet (B1), horizontal plane of unequal-energy  $e^+e^-$  collider PEP-II. (Courtesy M. Sullivan)

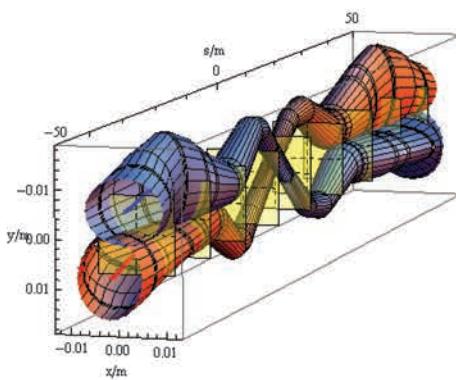


Figure 3:  $7\sigma$  beam envelopes in 2D separation around ALICE IP in equal-energy pp-collider LHC at 450 GeV; yellow planes are beam-beam encounters; horizontal separation at IP for injection, vertical separation at other encounters from crossing angle.

magnets) require detailed evaluation of changes to the optical functions, tunes, etc. Moreover these may differ between beams. E.g., in a pretzel scheme with horizontal separation  $x_p(s)$  in a FODO lattice the perturbation of the dispersion function  $D_{x0} \rightarrow D_x = D_{x0} + D_{x1}$  gives rise to a number of effects including a change in the longitudinal damping partition number (Sec.3.1.4) for  $e^+e^-$  beams,

$$J_s \approx \frac{dU(\delta)}{d\delta} = J_{s0} + \frac{2}{\mathcal{I}_2} \int K^2 x_p D_{x1} ds \quad (3)$$

where  $K$  is the quad gradient.

### 2.5.4.2 Long-range beam-beam effects

**Beam-beam kick** Consider the beam-beam force on test particle (or ion) (mass  $m$ , charge  $Ze$ ) at position  $\vec{x} = (x, y, z)$  due to the charge distribution,  $\rho(\vec{x})$ , of the opposing beam. (The coherent behavior of beams must be obtained by summing these forces over the beam and is not amenable to straightforward analytical treatment.) The potential [7] is

$$\begin{aligned} \Phi(\vec{x}) &= \frac{-1}{4\pi\epsilon_0} \int \frac{\rho(\vec{x}')}{|\vec{x} - \vec{x}'|} d\vec{x}' \\ &= \frac{-1}{4\pi\epsilon_0\sqrt{\pi}} \int_0^\infty \frac{dt}{t^{3/2}} \int \rho(\vec{x}') e^{-\frac{|\vec{x}-\vec{x}'|^2}{t}} d\vec{x}' \\ &= \frac{-Ne}{4\pi\epsilon_0\sqrt{\pi}} \int_0^\infty dt \\ &\times \frac{\exp\left(-\frac{x^2}{2\sigma_x^2+t} - \frac{y^2}{2\sigma_y^2+t} - \frac{z^2}{2\sigma_z^2+t}\right)}{\sqrt{(2\sigma_x^2+t)(2\sigma_y^2+t)(2\sigma_z^2+t)}} \end{aligned} \quad (4)$$

where the last form holds when  $\rho(\vec{x})$  is a 3-D gaussian with rms  $\sigma_{x,y,z}$  and normalized to  $Ne$ .

In the short bunch limit, the components of the beam-beam kick

$$\Delta p_{x,y} = -\frac{Ze}{p_0 c} \partial_{x,y} \int_{-\infty}^\infty \Phi(\vec{x}) dz \quad (5)$$

are conveniently computed in complex form [8, 9]  $i\Delta p_x + \Delta p_y = -\frac{ZNe}{\gamma} \Pi(x, y, \sigma_x, \sigma_y)$ , where  $r_0 = e^2/(4\pi\epsilon_0 mc^2)$ ,  $\Pi(x, y, a, b) =$

$$\begin{cases} \sqrt{\frac{2\pi}{a^2-b^2}} \left\{ W \left[ \frac{x+iy}{\sqrt{2(a^2-b^2)}} \right] \right. \\ \left. - e^{-\frac{x^2}{2a^2}-\frac{y^2}{2b^2}} W \left[ \frac{\frac{xb}{a}+i\frac{ya}{b}}{\sqrt{2(a^2-b^2)}} \right] \right\}, & (a>b, y\geq 0) \\ -\Pi^*(x, -y, a, b), & (a>b, y<0) \\ 2 \left( \frac{ix+y}{x^2+y^2} \right) \left( 1 - e^{-\frac{x^2+y^2}{a^2}} \right), & (a=b) \\ i\Pi^*(y, x, b, a), & (a<b) \end{cases} \quad (6)$$

[See also Sec.2.5.1.] Additional terms must be added [10] to make the full 6-D map symplectic. For non-Gaussian charge distributions see [11]. See [12] and Sec.2.5.1.2 for cases where the beam is thick longitudinally. Figs.4-5 show the behavior of the beam-beam kick.

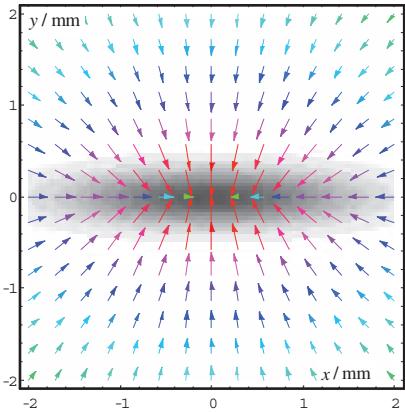


Figure 4: Kick vector field where the length (and color) of the arrows indicates the kick strength. Gray shading indicates the charge density. Same parameters as Fig.5.

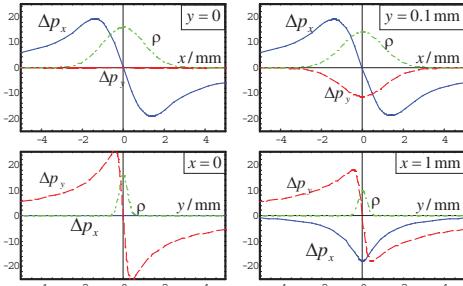


Figure 5: Beam-beam kicks in  $\mu\text{rad}$  felt by a 1 GeV,  $Z = -1$ , particle (e.g., electron) as a function of horizontal (left) and vertical (right) separations around a bunch with  $N = 10^{10}$ ,  $\sigma_x = 1 \text{ mm}$ ,  $\sigma_y = 0.2 \text{ mm}$  (e.g., positron bunch). In the plots on the left, the other separation is zero, while on the right a non-zero separation is included. Each plot has the same scales and the (scaled) Gaussian charge density is also shown.

At large separations,  $\min(x, y) \gg \max(\sigma_x, \sigma_y)$ , the beam looks like a point charge, the kick is purely radial and Eq.(6) can be approximated as

$$(\Delta p_x, \Delta p_y) = \frac{2ZNr_0}{\gamma} \frac{(x, y)}{x^2 + y^2} \quad (7)$$

This formula is useful in preliminary design of a separation scheme.

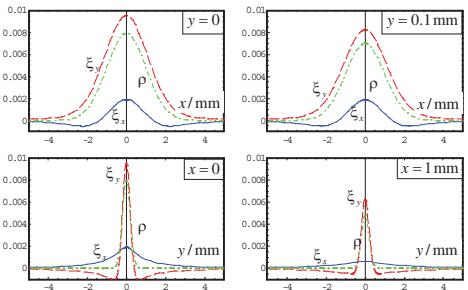


Figure 6: Parasitic beam-beam tune-shifts for a particle on the closed orbit, calculated perturbatively, for the same conditions as Fig.5 with  $\beta_x = \beta_y = 1 \text{ m}$ . The charge density is also shown.

**Beam-beam tune-shift** The beam-beam interaction produces tune-shifts

$$\xi_{x,y} = -\frac{\beta_{x,y}}{4\pi} \partial_{x,y} \Delta p_{x,y} \quad (8)$$

where  $(x, y)$  is the position of the closed orbit (center) of the weak beam relative to the opposing beam. In the case of a Gaussian beam, the change of variables  $t = u/(1-u)$  transforms Eq.(4) into a form convenient for numerical evaluation (Gaussian case can be evaluated analytically from Eq.(6)). Usually it is best to integrate separately over the neighborhoods of  $u = 0$  and  $u = 1$ . Examples are shown in Fig.6. For large separations, Eqs.(7) and (8) give

$$(\xi_x, \xi_y) = \frac{ZNr_0}{2\pi\gamma} \frac{(\beta_x, -\beta_y)(x^2 - y^2)}{(x^2 + y^2)^2} \quad (9)$$

The distribution of tune-shifts with betatron amplitudes inside a bunch can be derived analytically, extending Eqs.(21-22), Sec.2.5.1 in the case of a single encounter [14, 15]. Beyond this, simulation techniques [13, 16, 5] are generally more practical.

**Systems of many bunches** Since different bunches of each beam may see different sequences of encounters Eq.(2), there may be different orbits, optical functions, tunes and chromaticities for each bunch. Bunch pairs of bunches may see different separations, sizes and optical functions at the same encounter point. In practice, a self-consistent solution for the closed orbit and optics of a two-beam many-bunch system has to be found [19, 2] by simulation.

A *weak-strong* [17, 18, 2, 13] simulation treats a single test particle in the fields of an opposing beam, regarded as a fixed charge distribution. Beam-beam encounters are simulated as

special optical elements, using the kicks (6) with fixed  $\rho(\vec{x})$  with maps to propagate the orbit and optics between beam-beam encounters. The optics and tune of the test particle can be found as functions of its amplitude with respect to the closed orbit, giving *tune footprints* for the weak beam. A *strong-strong* simulation [5] computes mutual effects of the two beams. Some bunches in the system may be unstable and be lost, changing the stability of others which may subsequently be lost (e.g. the “PACMAN” effect) [20].

### Simple Criteria for adequate separation

Present understanding provides no universal criteria for the adequacy (beam lifetime, tolerable effects on beam sizes, etc.) of a separation scheme. Simple criteria [1] can be used in a design phase, pending thorough simulation, analysis or experiment. If  $X_i$  is the separation at encounter  $i$  (head-on excluded),  $\sigma_i$  a relevant beam size (e.g.  $\max(\sigma_x, \sigma_y)$ ), then some of the simplest criteria are

- $X_i > n_\sigma \sigma_i$ , where  $n_\sigma \approx 5.5\text{--}10$
- $\xi_{xi}, \xi_{yi} < \xi_{\max}$ , where  $\xi_{\max} \approx 10^{-4}\text{--}10^{-3}$
- $\sum_i \xi_{xi}, \sum_i \xi_{yi} < \xi_{\max}$ , where  $\xi_{\max} \approx 10^{-3}$

Applicability depends on the nature of the beam (generally stricter for protons than electrons) and other conditions. Experiments have shown that none of these criteria is strictly necessary or sufficient [3]. In hadron colliders, a strategy of minimizing the tune-spread in all the bunches is often followed [16, 3], possibly involving the compensation of tune-shifts between different parasitic encounters. Parasitic encounters may also induce amplitude-dependent chromaticity [14], drive additional resonances and diffusion, coherent effects, affect dynamic aperture, lifetime, combine with impedance effects, etc [21, 22, 24, 23].

### References

- [1] A.B. Temnykh et al, PAC 93, p.2007
- [2] LHC99, CERN-SL-99-039 AP (1999)
- [3] V. Shiltsev et al, PRST-AB 8 (2005) 101001
- [4] M. Sullivan et al, EPAC 04 MOPLT141 (2004)
- [5] M. Tawada et al, PAC 05 TPPP006 (2005)
- [6] R. Bailey et al, PAC 93, p.2013
- [7] K. Takayama, Lett. Al Nuovo Cimento 34 (1982) 190
- [8] M. Bassetti, G.A. Erskine, CERN-ISR-TH/80-06 (1980)
- [9] R. Talman, AIP Proc. 153 (1987) p.827
- [10] K. Hirata, H. Moshammer, F. Ruggiero, PA 40 (1993) 205
- [11] V. Ziemann, SLAC-PUB-5582 (1991)
- [12] J. Irwin et al, PAC 95, p.2904
- [13] D. Shatilov, PA 52 (1996) 65
- [14] T. Sen et al, PRST-AB 7 (2004) 041001
- [15] M. Meddahi, R. Schmidt, CERN SL/90-15 (1990)
- [16] W. Herr, PA 50 (1995) 69
- [17] Proc. 7th ICFA Beam Dynamics Workshop (1995), JINR Report 1996
- [18] A. Temnykh, D. Sagan, PAC 97, p.1768
- [19] E. Keil, CERN SL/95-75 (1995)
- [20] D. Neuffer, S. Peggs, SSC-63 (1986)
- [21] D. Siergiej et al, PRE 55 (1997)
- [22] Y. Papaphilippou et al, PRST-AB 5 (2002) 074001
- [23] E.G. Stern et al, PSRT-AB 13 (2010) 024401
- [24] M.A. Furman et al PAC 85, p.2297

## 2.6 POLARIZATION

### 2.6.1 Thomas-BMT Equation

*T. Roser, BNL*

Precession of polarization vector  $\vec{P}$  of a particle with mass  $m$  and charge  $Ze$  is given by [1]-[4]

$$\frac{d\vec{P}}{dt} = \vec{\Omega}_0 \times \vec{P} \quad (1)$$

$$\vec{\Omega}_0 = -\frac{Ze}{m\gamma} \left[ (1+G\gamma)\vec{B}_\perp + (1+G)\vec{B}_\parallel \right. \\ \left. + \left( G\gamma + \frac{\gamma}{\gamma+1} \right) \frac{\vec{E} \times \vec{v}}{c^2} \right]$$

$\vec{P}$  is defined in the particle rest frame,  $\vec{E}$  and  $\vec{B}$  in the laboratory frame.  $\vec{B} = \vec{B}_\perp + \vec{B}_\parallel$ ,  $\vec{B}_\parallel = (\vec{v} \cdot \vec{B})\vec{v}/v^2$ .

It is instructive to compare with the Lorentz force equation ( $\vec{E}_\parallel = 0$ )

$$\frac{d\vec{v}}{dt} = \vec{\Omega}_c \times \vec{v} \\ \vec{\Omega}_c = -\frac{Ze}{m\gamma} \left( \vec{B}_\perp + \frac{\gamma^2}{\gamma^2-1} \frac{\vec{E} \times \vec{v}}{c^2} \right) \quad (2)$$

In a frame rotating with  $\vec{v}$ :

for  $\tilde{\vec{B}}_\perp \neq 0$  :

$$\frac{d\vec{P}}{dt} = -\frac{Ze}{m\gamma} \left[ G\gamma \vec{B}_\perp \times \vec{P} \right] \quad (3)$$

for  $\tilde{\vec{E}}_\perp \neq 0$  :

$$\frac{d\vec{P}}{dt} = -\frac{Ze}{m\gamma} \left[ \left( G\gamma - \frac{\gamma}{\gamma^2-1} \right) \frac{\vec{E} \times \vec{v}}{c^2} \times \vec{P} \right]$$

$G = \frac{g-2}{2}$  is the anomalous magnetic moment;  $g = \frac{2m\mu}{ZeJ}$  is the gyromagnetic ratio.

|               | $G$        | $\frac{mc^2}{G}$ | [GeV] | $\left  \frac{ZeG}{mc} \right  \left[ \frac{\text{rad}}{\text{Tm}} \right]$ |
|---------------|------------|------------------|-------|---|
| e             | 0.00115965 | 0.440649         |       | 0.680342  |
| $\mu$         | 0.00116592 | 90.6220          |       | 0.00330816  |
| p             | 1.79285    | 0.523341         |       | 0.572843  |
| d             | -0.142987  | 13.1173          |       | 0.0228547   |
| $^3\text{He}$ | -4.18396   | 0.671228         |       | 0.893266  |
| $^3\text{H}$  | 7.91819    | 0.354744         |       | 0.845096  |

Note:

(i) In the rest frame,  
 $\vec{\Omega}_0 = -\frac{Zeq}{2m}\vec{B} = -\frac{Ze}{m}(1+G)\vec{B}$ .

(ii) For  $\vec{B}_{\parallel} = 0$  and  $\vec{E}_{\perp} = 0$  or  $\gamma \gg 1$ , one has  $(\Omega_0 - \Omega_c)/\Omega_c = G\gamma$ . This is the *spin tune* in this special case.

(iii) For the negative hydrogen ion the magnetic moment is given by the proton and the charge by the extra electron. For  $\vec{B}_{\parallel} = 0$  and  $\vec{E}_{\perp} = 0$ , one therefore has  $(\Omega_0 - \Omega_c)/\Omega_c = -(2+G\gamma)$ .

(iv) The “magic” energy for which  $\underline{\Omega_0 - \Omega_c}$  is independent of  $\vec{E}$  is given by  $\gamma = \sqrt{1 + \frac{1}{G}}$ .

## References

- [1] B.W. Montague, Phys. Rep. 113 (1984) 1
- [2] L.H. Thomas, Phil. Mag. 3 (1927) 1
- [3] V. Bargmann, L. Michel, V.L. Telegdi, PRL 2 (1959) 435
- [4] S.Y. Lee, Spin Dynamics and Snakes in Synchrotrons, World Scientific (1997)

### 2.6.2 Spinor Algebra

T. Roser, BNL

The coordinate frame is specified by the indices: (1,2,3) = (radial outward, longitudinal forward, vertical up) = ( $\hat{x}, \hat{s}, \hat{y}$ ). The Pauli matrices

$$\vec{\sigma} = (\sigma_1, \sigma_2, \sigma_3) = \begin{pmatrix} [0 & 1] \\ [1 & 0] \end{pmatrix}, \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

satisfy the following properties:

$$\begin{aligned} \sigma_1\sigma_1 &= \sigma_2\sigma_2 = \sigma_3\sigma_3 = I \\ \sigma_3\sigma_1 &= -\sigma_1\sigma_3 = i\sigma_2 \quad (\text{cyclic perm.}) \\ \text{tr}(\sigma_i) &= 0, \quad \det(\sigma_i) = -1 \\ (\vec{\sigma} \cdot \vec{a})(\vec{\sigma} \cdot \vec{b}) &= (\vec{a} \cdot \vec{b}) + i\vec{\sigma} \cdot (\vec{a} \times \vec{b}) \end{aligned} \quad (1)$$

The normalized vector  $\vec{P}$  has spinor representation

$$\begin{aligned} \vec{P} &= \psi^\dagger \vec{\sigma} \psi \\ \leftrightarrow \psi &= \frac{1}{\sqrt{2(P_3 + 1)}} \begin{bmatrix} 1 + P_3 \\ P_1 + iP_2 \end{bmatrix} \end{aligned} \quad (2)$$

Thomas-BMT equation can be written in the spinor or unitary representation,

$$\begin{aligned} \frac{d\vec{P}}{dt} &= \vec{\Omega} \times \vec{P} \\ \rightarrow \frac{d\psi}{dt} &= -\frac{i}{2} (\vec{\sigma} \cdot \vec{\Omega}) \psi \\ \text{or } \frac{d(\vec{\sigma} \cdot \vec{P})}{dt} &= -\frac{i}{2} [(\vec{\sigma} \cdot \vec{\Omega}), (\vec{\sigma} \cdot \vec{P})] \end{aligned} \quad (3)$$

Solution for constant  $\vec{\Omega}$  (Axis  $\hat{n}$ ,  $|\vec{\Omega}| = \omega$ ) is

$$\psi(t) = M(\hat{n}, \omega t) \psi(0), \quad \text{or} \quad (4)$$

$$(\vec{\sigma} \cdot \vec{P}(t)) = M(\hat{n}, \omega t) (\vec{\sigma} \cdot \vec{P}(0)) M^\dagger(\hat{n}, \omega t)$$

where the rotation operation  $M$  (angle  $\varphi = \omega t$  around axis  $\hat{n}$ ) is

$$\begin{aligned} M(\hat{n}, \varphi) &= \exp \left[ -i (\vec{\sigma} \cdot \hat{n}) \frac{\varphi}{2} \right] \\ &= \cos \left( \frac{\varphi}{2} \right) - i (\vec{\sigma} \cdot \hat{n}) \sin \left( \frac{\varphi}{2} \right) \end{aligned} \quad (5)$$

Conversely, given the rotation operator  $M$ , the rotation angle and axis can be found by the inverse operation,

$$\cos \left( \frac{\varphi}{2} \right) = \frac{1}{2} \text{tr}(M) \quad (6)$$

$$\hat{n} = \frac{i}{2 \sin \left( \frac{\varphi}{2} \right)} \text{tr}(\vec{\sigma} M) \quad (7)$$

The net spin rotation for one turn is given by the one-turn matrix

$$M_0(\theta) = M_n \cdots M_2 M_1 \quad (8)$$

where  $\theta$  is the starting (and ending) azimuth. The spin tune  $\nu_{sp}$  and the spin closed orbit  $\hat{n}_0$  (also called  $\hat{n}_0$  axis or stable spin direction) are

$$\cos(\pi\nu_{sp}) = \frac{1}{2} \text{tr}(M_0(\theta)) \quad (9)$$

independent of  $\theta$

$$\hat{n}_0(\theta) = \frac{i/2}{\sin(\pi\nu_{sp})} \text{tr}(\vec{\sigma} M_0(\theta)) \quad (10)$$

With a preferred axis of symmetry, usually vertical,

$$\begin{aligned} \sigma_+ &= \frac{1}{2} (\sigma_1 + i\sigma_2) = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \\ \sigma_- &= \frac{1}{2} (\sigma_1 - i\sigma_2) = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \end{aligned} \quad (11)$$

are a more useful representation. These matrices obey the following relations:

$$\begin{aligned} \sigma_+ \sigma_+ &= \sigma_- \sigma_- = 0 \\ \sigma_{\pm} \sigma_3 &= \mp \sigma_{\pm} \\ \sigma_3 \sigma_{\pm} &= \pm \sigma_{\pm} \end{aligned} \quad (12)$$

and the following transformations:

$$\begin{aligned}\exp\left(i\frac{\varphi}{2}\sigma_3\right)\sigma_{\pm} &= \exp\left(\pm i\frac{\varphi}{2}\right)\sigma_{\pm} \\ \sigma_{\pm}\exp\left(i\frac{\varphi}{2}\sigma_3\right) &= \exp\left(\mp i\frac{\varphi}{2}\right)\sigma_{\pm} \\ \exp\left(i\frac{\varphi}{2}\sigma_3\right)\sigma_{\pm} \quad \exp\left(-i\frac{\varphi}{2}\sigma_3\right) &= e^{\pm i\varphi}\sigma_{\pm}\end{aligned}\quad (13)$$

### 2.6.3 Spin Rotators and Siberian Snakes

T. Roser, BNL

“Spin Rotators” are devices that rotate  $\vec{P}$ , preferably without changing  $\vec{v}$ .

Examples:

Wien Filter: Transverse  $E_x$  and  $B_y$  with condition  $\frac{\vec{E} \times \vec{v}}{c^2} = \frac{1-\gamma^2}{\gamma^2} \vec{B}_{\perp}$ , the spin rotation is described by

$$\begin{aligned}\varphi &= \frac{Ze(1+G)}{mc\beta\gamma^2} \int B_{\perp} ds \\ M_{\text{Wien}} &= \cos \frac{\varphi}{2} - i\sigma_3 \sin \frac{\varphi}{2}\end{aligned}\quad (1)$$

Solenoid:

$$\begin{aligned}\varphi &= \frac{Ze(1+G)}{mc\beta\gamma} \int B_{\parallel} ds \\ M_{\text{Solenoid}} &= \cos \frac{\varphi}{2} - i\sigma_2 \sin \frac{\varphi}{2}\end{aligned}\quad (2)$$

Example:  $\varphi = 90^\circ$  and  $p = 1$  GeV/c requires  $\int B_{\parallel} ds = 1.88$  T-m for protons and  $\int B_{\parallel} ds = 5.23$  T-m for electrons.

Dipole: In lab frame:

$$\varphi = \frac{Ze(1+G\gamma)}{mc\beta\gamma} \int B_y ds$$

In frame rotating with  $\vec{v}$ :

$$\begin{aligned}\varphi &= \frac{ZeG}{mc\beta} \int B_y ds \\ M_{\text{Dipole}} &= \cos \frac{\varphi}{2} - i\sigma_3 \sin \frac{\varphi}{2}\end{aligned}\quad (3)$$

Example:  $\varphi = 90^\circ$  and  $\beta \approx 1$  requires  $\int B_y ds = 2.74$  T-m for protons and  $\int B_y ds = 2.31$  T-m for electrons.

Full twist helical dipole: [1]

$$\vec{B}(s) = B_0 \left( \sin \frac{2\pi s}{\lambda}, 0, \cos \frac{2\pi s}{\lambda} \right), \lambda > s > 0:$$

$$\begin{aligned}\varphi &= 2\pi \left[ \sqrt{1 + \chi^2} - 1 \right] \\ M &= \cos \frac{\varphi}{2} - i(\sigma_2 + \chi\sigma_3) \sin \frac{\varphi}{2} \\ \chi &\equiv \left( 1 + \frac{1}{G\gamma} \right) \frac{ZeG}{mc} \frac{B_0\lambda}{2\pi}\end{aligned}\quad (4)$$

Full Siberian snake: rotates  $\vec{P}$  by  $180^\circ$  ( $\varphi = \pi$ ) around an axis in the horizontal plane with angle  $\alpha$  from  $\hat{x}$  (Snake axis angle) [2]:

$$M_{\text{Snake}} = -i(\sigma_1 \cos \alpha + \sigma_2 \sin \alpha) \quad (5)$$

Type 1 snake: snake axis is long. ( $\alpha = 90^\circ$ );

Type 2 snake: snake axis is radial ( $\alpha = 0^\circ$ ). Note:

$$M_{\text{Dipole}} M_{\text{Snake}} M_{\text{Dipole}} = M_{\text{Snake}}.$$

Examples of one-turn matrices for rings:

1. Ideal ring without spin rotators or snakes:

$$M_0(\theta) = \exp(-i\sigma_3\pi G\gamma), \quad \nu_{\text{sp}} = G\gamma \quad (6)$$

2. Figure-8 ring without spin rotators or snakes:

$$M_0(\theta) = 1, \quad \nu_{\text{sp}} = 0 \quad (7)$$

3. Ring with solenoid (partial type 1 Siberian snake) at  $\theta_0 = 0$  and spin rotation  $\varphi$  [4]:

$$\begin{aligned}M_0(\theta) &= \cos \frac{\varphi}{2} \cos (\pi G\gamma) \\ &- i\sigma_1 \sin \frac{\varphi}{2} \sin ((\pi - \theta) G\gamma) \\ &- i\sigma_2 \sin \frac{\varphi}{2} \cos ((\pi - \theta) G\gamma) \\ &- i\sigma_3 \cos \frac{\varphi}{2} \sin (\pi G\gamma) \\ \nu_{\text{sp}} &= \frac{1}{\pi} \cos^{-1} \left[ \cos \frac{\varphi}{2} \cos (\pi G\gamma) \right]\end{aligned}\quad (8)$$

4. Ring with full Siberian snake with axis angle  $\alpha$  at  $\theta_0 = 0$ :

$$\begin{aligned}M_0(\theta) &= -i[\sigma_1 \cos(\alpha - (\pi - \theta)G\gamma) \\ &+ \sigma_2 \sin(\alpha - (\pi - \theta)G\gamma)]\end{aligned}\quad (9)$$

$$\nu_{\text{sp}} = \frac{1}{2}$$

Note: with a type 1 snake ( $\alpha = 90^\circ$ ) at  $\theta = 0$ ,  $\hat{n}_0(\theta = \pi) = \hat{s}$ , i.e. the beam polarization is longitudinal.

5. Ring with two full Siberian snakes with axis angles  $\alpha_a$  and  $\alpha_b$  at  $\theta_a$  and  $\theta_b$ :

$$M_0(\theta) = -\exp(-i\sigma_3\chi) \quad (10)$$

$$\chi \equiv \alpha_b - \alpha_a + (\pi - \theta_b + \theta_a)G\gamma$$

Note:  $M_0$  is energy independent for  $\theta_b - \theta_a = \pi$ ; then  $\nu_{\text{sp}} = \frac{1}{\pi}(\alpha_b - \alpha_a)$ . For  $(\alpha_b - \alpha_a) = \frac{\pi}{2}$  then  $\nu_{\text{sp}} = \frac{1}{2}$ .

## Sec.2.6: POLARIZATION

6. Ring with  $N$  pairs of full Siberian snakes with axis angles  $\alpha_a^i$  and  $\alpha_b^i$  at  $\theta_a^i$  and  $\theta_b^i$ :

$$M_0(\theta) = (-)^N \exp(-i\sigma_3\chi) \quad (11)$$

where  $\chi = \sum_{i=1}^N (\alpha_b^i - \alpha_a^i) + [\pi - \sum_{i=1}^N (\theta_b^i - \theta_a^i)] G\gamma$ .

Note:  $M_0$  is energy independent for  $\sum_{i=1}^N (\theta_b^i - \theta_a^i) = \pi$ ; then  $\nu_{sp} = \frac{1}{\pi} \sum_{i=1}^N (\alpha_b^i - \alpha_a^i)$ . There are many parameter sets that result in  $\nu_{sp} = \frac{1}{2}$ . See [5, 6] for parameter sets that minimize the amplitude dependence of the spin tune.

## References

- [1] V. Ptitsin, Y. Shatunov, NIM A398 (1997) 126
- [2] Ya.S. Derbenev, A.M. Kondratenko, PA 8 (1978) 115
- [3] Operating example at HERA, D.P. Barber et al, Phys. Lett. 343B (1995) 436
- [4] T. Roser, Proc. Workshop on Siberian Snakes and Depolarizing Techniques (1989) p.1442
- [5] G.H. Hoffstatter, High-Energy Polarized Proton Beams, Springer (2006)
- [6] S.R. Mane, NIM A587 (2008) 188

### 2.6.4 Depolarizing Resonances and Spin Flippers

*T. Roser, BNL*

Thomas-BMT equation with azimuthal coordinate  $\theta$  as independent variable and the fields expressed in terms of the particle coordinates:

$$\begin{aligned} \frac{d\psi}{d\theta} &= -\frac{i}{2} \begin{bmatrix} G\gamma & -\xi \\ -\xi^* & -G\gamma \end{bmatrix} \psi \quad (1) \\ \xi &= -\rho y'' (1 + G\gamma) \\ &- i \left[ (1 + G\gamma) y' - \rho (1 + G) \left( \frac{y}{\rho} \right)' \right] \end{aligned}$$

where  $\rho$  is the bending radius. Resonance strength is [1]

$$\epsilon_K = \frac{1}{2\pi} \oint \xi \exp(iK\theta) d\theta \quad (2)$$

where the loop integral keeps only the  $e^{-iK\theta}$  frequency component in  $\xi$ . The case  $K = kP \pm \nu_y$  gives intrinsic resonance driven by vertical betatron motion;  $P$  is super periodicity. The case  $K = k$  gives imperfection resonance driven by vertical closed orbit distortions. One way to reduce resonance strength is in [2].

For an isolated resonance,  $\xi = \epsilon_K \exp(-iK\theta)$ . In a frame rotating around  $\hat{y}$  with tune  $K$ ,

$$\begin{aligned} \psi_K &= \exp\left(\frac{i}{2}K\theta\sigma_3\right)\psi \\ \frac{d\psi_K}{d\theta} &= -\frac{i}{2} \begin{bmatrix} G\gamma - K & -\epsilon_K \\ -\epsilon_K^* & K - G\gamma \end{bmatrix} \psi_K \quad (3) \end{aligned}$$

with solution (for small betatron amplitudes,  $\hat{n}_0$  and  $\nu_{sp}$  are evaluated along the closed-orbit),

$$\begin{aligned} \psi_K(\theta) &= \exp\left(-\frac{i}{2}\nu_{sp}\theta(\hat{n}_0 \cdot \vec{\sigma})\right)\psi_K(0) \\ \nu_{sp} &= \sqrt{(G\gamma - K)^2 + |\epsilon_K|^2} \quad (4) \\ (\hat{n}_0 \cdot \vec{\sigma}) &= \frac{[(G\gamma - K)\sigma_3 - \epsilon_K\sigma_+ - \epsilon_K^*\sigma_-]}{\nu_{sp}} \end{aligned}$$

Solutions for a single isolated resonance in a ring with snakes also exist, but more complicated [3].

Passage through an isolated resonance with a linear crossing pattern is described by the Froissart-Stora Equation [4],

$$\frac{P_{final}}{P_{initial}} = 2 \exp\left(-\frac{\pi|\epsilon_K|^2}{2\alpha}\right) - 1 \quad (5)$$

with  $\alpha = \frac{d(G\gamma)}{d\theta}$  (crossing speed). Fast passage  $\rightarrow P_{final} \approx P_{initial}$ . Slow passage  $\rightarrow P_{final} \approx -P_{initial} \rightarrow$  spin flip. Passage with a piecewise linear zigzag pattern is described in [5].

Artificial resonance from local oscillating longitudinal field [6, 7] ( $\omega$  = applied frequency,  $\omega_0$  = revolution frequency):

$$\begin{aligned} B_{||} &= \hat{B}_{||} \cos(\omega t) \\ \rightarrow \epsilon_K &= \frac{(1+G)}{4\pi} \frac{\int \hat{B}_{||} ds}{B\rho}; \quad K = n + \frac{\omega}{\omega_0} \quad (6) \end{aligned}$$

Spin flip by ramping artificial resonance through resonance condition with speed  $\alpha$ :

$$\alpha = \frac{K_{end} - K_{start}}{2\pi N} \quad (7)$$

where  $N$  is number of turns during ramp. For more than 99% spin flip,

$$K_{end} - K_{start} \geq 14\epsilon_K; \quad \epsilon_K \geq \frac{1}{140N} \quad (8)$$

For local radial oscillating field (AC dipole) resonance strength is sum of direct contribution

$$\begin{aligned} B_{\perp} &= \hat{B}_{\perp} \cos(\omega t) \\ \rightarrow \epsilon_K &= \frac{(1+G\gamma)}{4\pi} \frac{\int \hat{B}_{\perp} ds}{B\rho}; \quad K = n + \frac{\omega}{\omega_0} \quad (9) \end{aligned}$$

plus contribution from oscillating vertical orbit motion in the lattice quadrupoles. The latter contribution is large near a strong intrinsic resonance [8] or in high energy rings with many quadrupoles [9].

In a ring with pairs of Siberian snakes additional higher order “snake” resonances [10] occur when the fractional vertical betatron tune

$$\Delta\nu_y = \pm \frac{\nu_{sp} - k}{(2l - 1)} = \frac{2k - 1}{2(2l - 1)} \text{ for } \nu_{sp} = \frac{1}{2} \quad (10)$$

With a single snake or vertical closed orbit distortions snake resonances also occur when

$$\Delta\nu_y = \pm \frac{\nu_{sp} - k}{(2l)} = \frac{2k - 1}{2(2l)} \text{ for } \nu_{sp} = \frac{1}{2} \quad (11)$$

Snake resonances are particularly strong at energies near strong intrinsic resonances of the ring without snakes.

## References

- [1] E.D. Courant, R. Ruth, BNL-51270 (1980)
- [2] A.W. Chao, Ya.S. Derbenev, PA 36 (1991) 25
- [3] S.R. Mane, NIM A498 (2003) 1
- [4] M. Froissart, R. Stora, NIM 7 (1960) 297
- [5] A.W. Chao, PRST-AB 8, 104001 (2005)
- [6] D.D. Caussyn et al, PRL 73 (1994) 2857
- [7] M. Bai et al, PRST-AB 8 (2005) 99001
- [8] M. Bai et al, PRL 80 (1998) 4673
- [9] V. Ptitsyn et al, NIM A608 (2009) 225
- [10] S.Y. Lee, S. Tepikian, PRL 56 (1986) 1635

### 2.6.5 Polarized Hadron Beams and Siberian Snakes

*A.D. Krisch, U. Michigan  
M.A. Leonova, FNAL*

In 1973, the first high energy polarized proton beam was successfully accelerated in the Argonne ZGS. The depolarizing resonances in the 12 GeV weak-focusing ZGS were fairly weak; thus, maintaining polarization only needed careful orbit control to overcome the imperfection resonances and moderately fast betatron tune jumps to overcome the intrinsic resonances [1]. Polarized protons were then successfully accelerated to several GeV at Saturne [2] and KEK [3]. The plan for higher energy polarized protons started at a 1977 workshop [4] two years prior to the 1979 ZGS shutdown.

Polarized protons were first accelerated at AGS in 1984. Maintaining the polarization was more difficult, since the 30 GeV strong-focusing

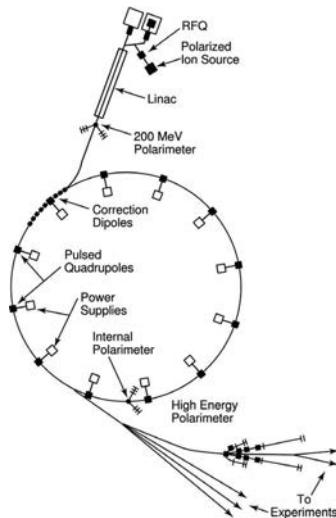


Figure 1: The individual depolarizing resonance correction hardware installed in AGS [5].

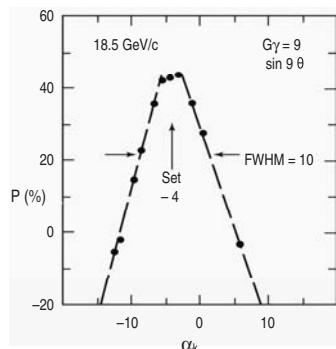


Figure 2: Individually correcting the  $G\gamma = 9$  imperfection resonance in AGS [5].

AGS has strong depolarizing resonances. As shown in Fig.1, AGS needed many complex hardware modifications. To maintain the polarization up to 22 GeV, 39 imperfection resonances were corrected individually by using 96 small dipoles to form the correct harmonic wave of horizontal magnetic field as each resonance was passed; moreover, 5 intrinsic resonances were jumped (only partly successfully) using 10 formidable 22 MW 1.6  $\mu$ s pulsed quadrupoles. Fig.2 shows a typical AGS imperfection resonance correction curve [5]. Each polarized beam tune-up required 3-7 weeks of dedicated AGS operation.

## Sec.2.6: POLARIZATION

This individual resonance correction technique was impractical at higher energy, since the number of imperfection resonances to be crossed is  $E/(0.52 \text{ GeV})$ . Thus, it became clear that the untested Siberian snakes [6] (Sects.2.6.3, 2.6.4) were needed [7] to accelerate polarized protons above 30 GeV.

The first Siberian snake tests were performed at IUCF's 500-MeV Cooler Ring. The snake was a 2 T-m superconducting solenoid installed in a 6-m straight section. Fig.3 shows data when crossing the  $G\gamma = 2$  imperfection resonance, with and without the snake. Without the snake, full polarization was achieved only when the imperfection magnetic fields were exactly corrected; any small imperfection field destroyed the polarization. With the snake, full polarization was maintained over the entire range of imperfection fields [8]. Fig.4 shows the snake also overcoming an intrinsic resonance [9].

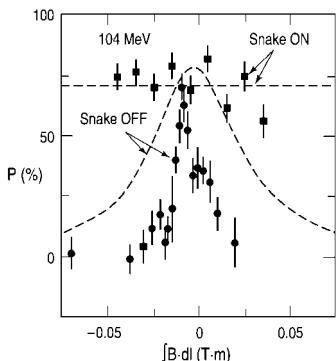


Figure 3: Siberian snake overcoming the  $G\gamma = 2$  imperfection resonance at IUCF [8].

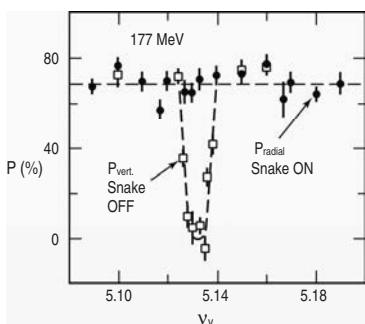


Figure 4: Siberian snake overcoming the strong intrinsic resonance  $G\gamma = 3 + \nu_y$  at IUCF [9].

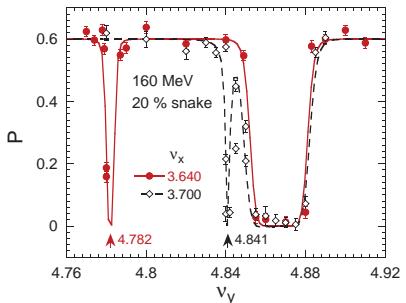


Figure 5: Observation of  $7 - \nu_y$  and  $1 + \nu_y - \nu_x$  spin resonances at IUCF [11].

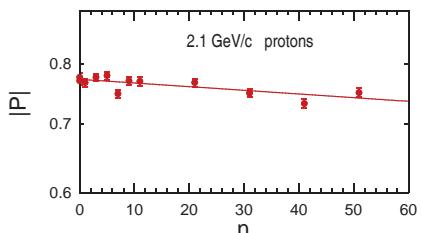


Figure 6: Multiple spin flipping at COSY [17, 14].

During 10 adiabatic ramps of a partial Siberian snake [10], there was no polarization loss when  $G\gamma$  was a half-integer ( $2\frac{1}{2}$ ), as suggested by Courant. Fig.5 shows the first firm evidence for a 2<sup>nd</sup>-order snake resonance ( $G\gamma = 1 + \nu_y - \nu_x$ ) [11]. Experiments were also performed at IUCF involving: synchrotron sideband resonances [9]; partial Siberian snakes [12]; snake resonances [13]; and rf-induced spin resonances, which can calibrate the beam energy or flip the spin direction with over 99.9% efficiency [14]. In 2002, the Cooler Ring's final year, striking and unexpected behavior was found [15] in the first spin manipulation of simultaneously vector- and tensor-polarized deuterons, which are spin-1 bosons.

After Cooler Ring's shutdown, its spin program was transferred to the COSY Cooler Synchrotron in Jülich. COSY had earlier accelerated polarized protons to 3 GeV using moderately fast betatron tune jumps [16].

The new SPIN@COSY team then demonstrated that one could also have proton spin-flip efficiencies as high as  $99.92 \pm 0.04\%$  [17], as shown in Fig.6, at energies 10 times that at IUCF [14]. They also confirmed [18], with better

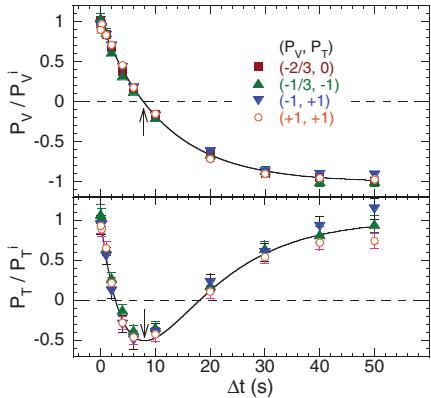


Figure 7: Spin-manipulating deuterons' vector and tensor polarizations at COSY [18, 15].

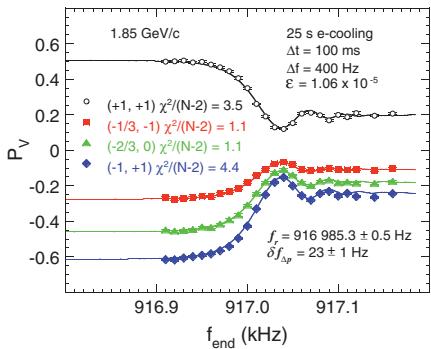


Figure 8: Measured deuteron vector polarizations vs. rf-solenoid end frequency, with CM fits [21].

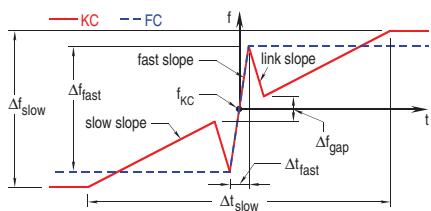


Figure 9: Kondratenko Crossing (KC) {solid line} and Fast Crossing (FC) {dashed} patterns [23].

precision, the striking behavior [15] of deuterons when their spins are manipulated; their tensor polarization's quantum mechanical behavior (Fig.7) may have some significance outside of accelerator physics.

The Chao matrix (CM) formalism [19] allows analytic calculations of a beam's polarization behavior inside a spin resonance, where the Froissart-Stora formula [20] is not applicable. The formalism's prediction of polarization oscillations when crossing an isolated spin resonance were tested at COSY [21]. The good fits to the experiment's precise data (Fig.8) demonstrate the formalism's ability to obtain experimental values for a spin resonance's position, its strength, and its frequency spread due to beam momentum spread.

Kondratenko et al [22] proposed a new technique for overcoming medium-strength depolarizing resonances when Siberian snakes are not practical and Fast Crossing (FC) is too weak. Each resonance is crossed using the Kondratenko Crossing (KC) pattern (Fig.9). KC was tested at COSY with deuterons using an rf-induced spin resonance [23]. KC reduced the depolarization far better than FC at the same fast crossing rate, by factors of  $4.7 \pm 0.3$  and  $19^{+12}_{-5}$  for unbunched and bunched beams, respectively.

The SPIN Collaboration also produced detailed plans for accelerating polarized protons at Fermilab (6 snakes in Tevatron and 2 in the 120-GeV Main Injector) [24-a], in the 920-GeV HERA proton ring (4 snakes) [24-b] and in the 20-TeV SSC (26 snakes in each 20-TeV ring) [24-c]. These projects were not implemented.

Brookhaven's RHIC-SPIN Collaboration installed 2 Siberian snakes in each RHIC ring, with 25 GeV/c polarized protons injected from AGS [25, 26] (Fig.10). With help from RIKEN in Japan, this 100 GeV/c on 100 GeV/c polarized p-p collider started operating in 2000.

In the mid-1990s, polarized beam acceleration again started at AGS to prepare for injecting polarized protons into RHIC. However, the AGS problems discussed earlier have remained; its 2-25 GeV/c range is too low for full Siberian snakes to be practical and too high for the resonance jumping techniques to work efficiently. To help solve this problem another workshop was held in 2002 [27]. As suggested by Roser [28], a cold partial (15%) snake, was installed in AGS. The result of using two partial helical dipole snakes [29] to overcome about 40 imperfection resonances in AGS is shown in Fig.11 [26]. The final AGS polarization during the 2006 run was about 65% [26, 29].

More importantly, RHIC has demonstrated that the Siberian snake technique worked well

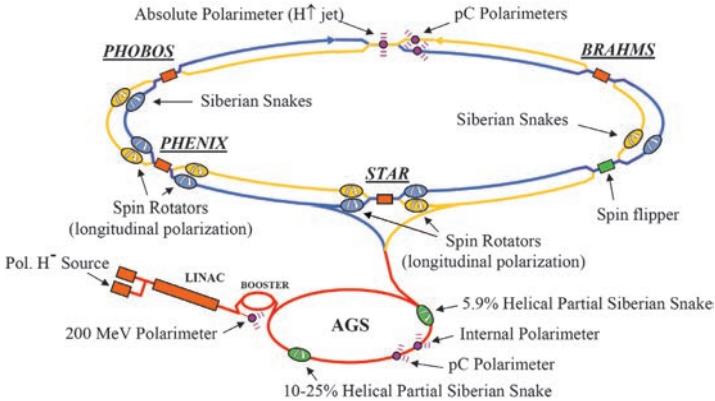


Figure 10: RHIC accelerator complex in 2008 with polarized proton elements highlighted [25-31].

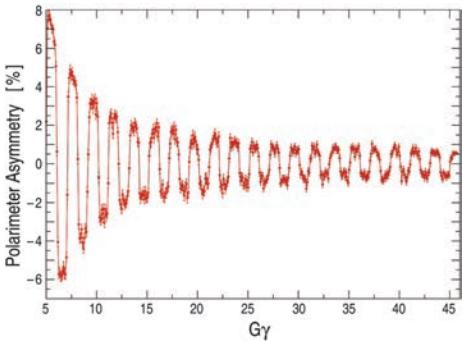
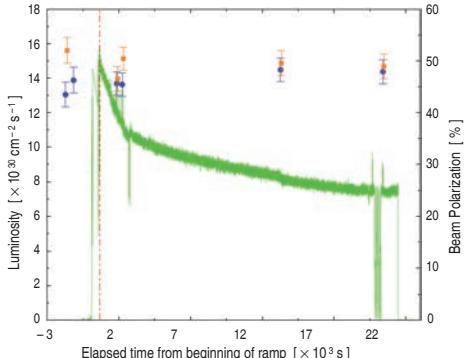


Figure 11: The measured asymmetry in AGS with two partial helical dipole snakes [26].

at high energy; when properly tuned, its beams were accelerated from 25 GeV/c at injection to 100 GeV/c and then stored for many hours with essentially no polarization loss, as shown in Fig.12 [30]. This allowed measurements of spin-spin effects in 100 GeV/c p-p scattering ( $\sqrt{s} = 200$  GeV). Moreover, in recent runs, a 45% polarization was reached in brief efforts to accelerate and store 250 GeV/c protons [26, 31].

## References

- [1] T.L. Khoi et al, PA 6 (1975) 213
- [2] J.L. Laclare et al, JP 46 (1985) C2-499
- [3] H. Sato et al, NIM A 272 (1988) 617
- [4] A.D. Krisch et al, AIP Conf. Proc. 42 (1978)
- [5] F.Z. Khiari et al, PRD 39 (1989) 45
- [6] Ya.S. Derbenev, A.M. Kondratenko, PA 8 (1978) 115
- [7] A.D. Krisch et al, AIP Conf. Proc. 145 (1986)
- [8] A.D. Krisch et al, PRL 63 (1989) 1137
- [9] J.E. Goodwin et al, PRL 64 (1990) 2779
- [10] R.A. Phelps et al, PRL 72 (1994) 1479
- [11] C. Ohmori et al, PRL 75 (1995) 1931
- [12] B.B. Blinov et al, PRL 73 (1994) 1621
- [13] R.A. Phelps et al, PRL 78 (1997) 2772
- [14] B.B. Blinov et al, PRL 88 (2002) 014801
- [15] V.S. Morozov et al, PRL 91 (2003) 214801
- [16] A. Lehrach et al, AIP Proc. 675 (2003) 153
- [17] M.A. Leonova et al, PRL 93 (2004) 224801
- [18] V.S. Morozov et al, PRST-AB 8 (2005) 061001
- [19] A.W. Chao, PRST-AB 8 (2005) 104001
- [20] M. Froissart, R. Stora, NIM 7 (1960) 297
- [21] V.S. Morozov et al, PRL 100 (2008) 054801
- [22] A.M. Kondratenko et al, Phys. of Particles and Nuclei Lett. 1 (2004) 266
- [23] V.S. Morozov et al, PRL 102 (2009) 244801

Figure 12: A typical polarized proton store in both RHIC rings during 2005 at  $\sqrt{s} = 200$  GeV [30].

- [24] -a SPIN Collaboration, U. Michigan Report UM HE 95-09 (1995); -b ibid. UM HE 96-20 (1996); -c ibid. UM HE SSC-LOI (1990)
- [25] T. Roser, PAC95, 3154
- [26] T. Roser, AIP Proc. 1149 (2009) 180; H. Huang et al, ibid. 767
- [27] A.D. Krisch et al, AIP Conf. Proc. 667 (2003)
- [28] T. Roser, ibid. 130
- [29] H. Huang et al, PRL 99 (2007) 154801
- [30] T. Roser, PAC05 358; H. Huang et al, ibid. 1404
- [31] M. Bai et al, PRL 96 (2006) 174801

## 2.6.6 Radiative Polarization in Electron Storage Rings

D.P. Barber, DESY  
G. Ripken, Deceased 2004

**Sokolov-Ternov (ST) effect [1]** Relativistic electrons in a storage ring emit synchrotron radiation. A very small fraction of the radiated photons cause spin flip. For electron spins aligned along a uniform magnetic field, the  $\uparrow\downarrow$  and  $\downarrow\uparrow$  flip rates differ and then spin polarization builds up antiparallel to the field. Positrons become polarized parallel to the field. The transition rates for electrons are

$$\begin{aligned} W_{\uparrow\downarrow} &= \frac{5\sqrt{3}}{16} \frac{r_e \gamma^5 \hbar}{m_e |\rho|^3} \left( 1 + \frac{8}{5\sqrt{3}} \right) \\ W_{\downarrow\uparrow} &= \frac{5\sqrt{3}}{16} \frac{r_e \gamma^5 \hbar}{m_e |\rho|^3} \left( 1 - \frac{8}{5\sqrt{3}} \right) \end{aligned} \quad (1)$$

For positrons, interchange plus and minus signs here and elsewhere.

The equilibrium polarization in a uniform magnetic field is independent of  $\gamma$ ,

$$P_{st} = \frac{W_{\uparrow\downarrow} - W_{\downarrow\uparrow}}{W_{\uparrow\downarrow} + W_{\downarrow\uparrow}} = \frac{8}{5\sqrt{3}} = 0.9238 \quad (2)$$

For a beam initially unpolarized, the time dependence for build-up to equilibrium is

$$P(t) = P_{st} [1 - \exp(-t/\tau_0)] \quad (3)$$

with build-up rate

$$\tau_0^{-1} = \frac{5\sqrt{3}}{8} \frac{r_e \gamma^5 \hbar}{m_e |\rho|^3} \quad (4)$$

The time  $\tau_0$  depends strongly on  $\gamma$  and  $\rho$  but is typically minutes or hours. In a flat ring in which all bending magnets have the same  $\rho$  just average (4) over the circumference  $C$ ,

$$\tau_0^{-1} [\text{s}^{-1}] \approx \frac{2\pi}{99} \frac{E[\text{GeV}]^5}{C[\text{m}] \rho[\text{m}]^2} \quad (5)$$

**Baier-Katkov flip rate** For electron spins initially aligned along an arbitrary unit vector  $\hat{\xi}$ , Eq.(1) generalizes to [2]

$$W = \frac{1}{2\tau_0} \left[ 1 - \frac{2}{9} (\hat{\xi} \cdot \hat{s})^2 + \frac{8}{5\sqrt{3}} \hat{\xi} \cdot \hat{b} \right] \quad (6)$$

where  $\hat{s}$  = direction of motion and  $\hat{b} = (\hat{s} \times \hat{\dot{s}})/|\hat{s}|$ .  $\hat{b}$  is the magnetic field direction if the electric field vanishes and the motion is perpendicular to the magnetic field. The corresponding instantaneous rate of polarization along  $\hat{\xi}$  is

$$\tau_{bk}^{-1} = \tau_0^{-1} \left[ 1 - \frac{2}{9} (\hat{\xi} \cdot \hat{s})^2 \right] \quad (7)$$

**Thomas-BMT equation** Neglecting radiative spin flip, the motion of the CM spin expectation value  $\vec{\xi}$  of a charged particle traveling in EM fields is governed by the Thomas-BMT equation  $d\vec{\xi}/dt = \vec{\Omega} \times \vec{\xi}$  (Sec.2.6.1).

We write

$$\vec{\Omega} = \vec{\Omega}^{co} + \vec{\omega}^{sb} \quad (8)$$

where  $\vec{\Omega}^{co}$  is due to the fields on the closed orbit, whence  $\vec{\Omega}^{co}(s+C) = \vec{\Omega}^{co}(s)$ .  $\vec{\Omega}^{co} = \vec{\Omega}^{ref} + \vec{\omega}^{imp}$ , where  $\vec{\Omega}^{ref}$  contains the design fields and  $\vec{\omega}^{imp}$  represents the effects of magnet misalignments, correction fields etc.  $\vec{\omega}^{sb}$  is due to 6-D synchro-betatron motion with respect to the closed orbit.

On the closed orbit the T-BMT equation

$$\frac{d}{dt} \vec{\xi} = \vec{\Omega}^{co} \times \vec{\xi} \quad (9)$$

can be solved in the form

$$\vec{\xi}(s) = \mathbf{R}_{3 \times 3}^{co}(s, s_0) \vec{\xi}(s_0) \quad (10)$$

where  $\mathbf{R}_{3 \times 3}^{co}$  is a rotation matrix. The real unit eigenvector (rot. axis) for the 1-turn matrix  $\mathbf{R}_{3 \times 3}^{co}(s+C, s)$ , denoted by  $\hat{n}_0(s)$ , is the periodic spin solution on the closed orbit. For a perfectly aligned flat ring with no solenoids,  $\hat{n}_0(s) = \pm \hat{y}$ . The 1-turn matrix has a complex conjugate pair of eigenvalues  $e^{\pm i2\pi\nu_{sp}}$ . Given  $\hat{n}_0$ , we define the unit vectors  $(\hat{m}_0, \hat{l}_0)$  such that  $\hat{m}_0 = \hat{l}_0 \times \hat{n}_0$  and  $\hat{l}_0 = \hat{n}_0 \times \hat{m}_0$  fulfill Eq.(9), and such that

$$\begin{aligned} &\hat{m}_0(s_0 + C) + i\hat{l}_0(s_0 + C) \\ &= e^{i2\pi\nu_{sp}} [\hat{m}_0(s_0) + i\hat{l}_0(s_0)] \end{aligned} \quad (11)$$

The  $(\hat{m}_0, \hat{l}_0)$  are usually not periodic in  $s$ . But with a further rotation by an angle  $\psi_{sp}(s)$  around  $\hat{n}_0$  we can construct the vectors  $(\hat{m}, \hat{l})$ ,

$$\hat{m}(s) + i\hat{l}(s) = e^{-i\psi_{sp}(s)} [\hat{m}_0(s) + i\hat{l}_0(s)] \quad (12)$$

By choosing  $\psi_{\text{sp}}(s + C) - \psi_{\text{sp}}(s) = 2\pi\nu_{\text{sp}}$ , the set  $(\hat{n}_0, \hat{m}, \hat{l})$  is then periodic in  $s$  with period  $C$ . The vectors  $(\hat{m}, \hat{l})$  are needed in Sec.2.6.7.

The closed orbit *spin tune*  $\nu_{\text{sp}}$  is the number of spin precessions per turn around  $\hat{n}_0$ . For a perfectly aligned flat ring without solenoids  $\nu_{\text{sp}} = a\gamma_0$ , where  $a = (g - 2)/2$  (Sec.2.6.1) and  $\gamma_0$  is the Lorentz factor for the beam energy.<sup>1</sup> Only the fractional part of  $\nu_{\text{sp}}$  can be extracted from the numerical values of the eigenvalues  $e^{\pm i2\pi\nu_{\text{sp}}}$ .

### Baier-Katkov-Strakhovenko (BKS) equation

Neglecting the effect of stochastic (synchrotron radiation) photon emission on the trajectory and imagining that all particles remain on the closed orbit, the equation of motion for electron polarization is [3]

$$\frac{d\vec{P}}{dt} = \vec{\Omega}^{\text{co}} \times \vec{P} - \frac{1}{\tau_0(s)} \left[ \vec{P} - \frac{2}{9} \hat{s} (\vec{P} \cdot \hat{s}) + \frac{8}{5\sqrt{3}} \hat{b}(s) \right]$$

In the case of horizontal motion in a vertical magnetic field, we have  $\vec{\Omega} = (a\gamma c/\rho)\hat{y}$ , and  $\hat{b}(s) = \hat{y}$ .

From the BKS equation, the generalized Sokolov-Ternov formula for the asymptotic electron polarization in arbitrary magnetic fields along the closed orbit is

$$\vec{P}_{\text{bks}} = -\frac{8}{5\sqrt{3}} \hat{n}_0 \frac{\oint ds \frac{\hat{n}_0(s) \cdot \hat{b}(s)}{|\rho(s)|^3}}{\oint ds \frac{[1 - \frac{2}{9}(\hat{n}_0(s) \cdot \hat{s})^2]}{|\rho(s)|^3}} \quad (13)$$

See [4] for time scales. Usually, in rings containing dipole spin rotators (Sec.2.6.3) the polarization  $|\vec{P}_{\text{bks}}|$  cannot reach 0.9238 [5].

The BKS polarization build-up rate is

$$\tau_{\text{bks}}^{-1} = \frac{5\sqrt{3} r_e \gamma^5 \hbar}{8 m_e} \frac{1}{C} \oint ds \frac{[1 - \frac{2}{9}(\hat{n}_0 \cdot \hat{s})^2]}{|\rho(s)|^3} \quad (14)$$

This is in accord with Eq.(7) by replacing  $\xi \rightarrow \hat{n}_0$  and averaging.

**Radiative depolarization** The stochastic element of photon emission together with damping determines the equilibrium phase-space density distribution. The same photon emission also imparts a stochastic element to  $\vec{\omega}^{\text{sb}}$  and then, via the T-BMT equation, spin diffusion (and thus depolarization) can occur [6]. The attainable polarization is the result of a balance between the Sokolov-Ternov effect and this radiative depolarization. In the approximation that the orbital motion is linear, the *value* of the polarization is essentially the same at each point in phase space and

<sup>1</sup> In this and the next sections, we use the symbol “ $a$ ” instead of “ $G$ ” used in the rest of the Handbook.

azimuth and the polarization is aligned along the Derbenev-Kondratenko vector  $\hat{n}$  [7, 8].

The unit vector field  $\hat{n}$  (the “*invariant spin field*”) [9, 10] depends on  $s$  and  $\vec{u} \equiv (x, p_x, y, p_y, z, \delta)$ .  $\hat{n}(\vec{u}; s)$  satisfies the T-BMT equation at  $(\vec{u}; s)$  and is periodic:  $\hat{n}(\vec{u}; s) = \hat{n}(\vec{u}; s + C)$ . On the closed orbit  $\hat{n}(\vec{u}; s)$  reduces to  $\hat{n}_0(s)$ .

**Derbenev-Kondratenko-Mane formula** Taking into account radiative depolarization due to photon-induced longitudinal recoils, the equilibrium electron polarization along the  $\hat{n}$  field is [7, 11, 8]

$$P_{\text{dk}} = \frac{-\frac{8}{5\sqrt{3}} \times \oint ds \left\langle \frac{1}{|\rho(s)|^3} \hat{b} \cdot (\hat{n} - \frac{\partial \hat{n}}{\partial \delta}) \right\rangle_s}{\oint ds \left\langle \frac{1}{|\rho(s)|^3} (1 - \frac{2}{9}(\hat{n} \cdot \hat{s})^2 + \frac{11}{18}(\frac{\partial \hat{n}}{\partial \delta})^2) \right\rangle_s} \quad (15)$$

where  $\langle \rangle_s$  denotes an average over phase space at azimuth  $s$ . This formula differs from Eq.(13) by the inclusion of the terms with  $\frac{\partial \hat{n}}{\partial \delta}$  and use of  $\hat{n}$  instead of  $\hat{n}_0$ . The ensemble average of the polarization is

$$\vec{P}_{\text{ens,dk}}(s) = P_{\text{dk}} \langle \hat{n} \rangle_s \quad (16)$$

and  $\langle \hat{n} \rangle_s$  is very nearly aligned along  $\hat{n}_0(s)$  (see the angle estimate below). The *value* of the ensemble average,  $P_{\text{ens,dk}}(s)$ , is essentially independent of  $s$ .

The effect of transverse recoil can also be included but contributes derivative terms analogous to  $\frac{\partial \hat{n}}{\partial \delta}$  which are typically a factor  $\gamma$  smaller than  $\frac{\partial \hat{n}}{\partial \delta}$  and can be neglected unless  $\frac{\partial \hat{n}}{\partial \delta}$  is very small [12, 13]. If  $\frac{\partial \hat{n}}{\partial \delta}$  were to vanish, a  $P_{\text{dk}}$  of 99.2 % could be reached [12, 13, 8].

In the presence of radiative depolarization Eq.(14) becomes

$$\tau_{\text{dk}}^{-1} = \frac{5\sqrt{3} r_e \gamma^5 \hbar}{8 m_e} \frac{1}{C} \times \oint ds \left\langle \frac{1 - \frac{2}{9}(\hat{n} \cdot \hat{s})^2 + \frac{11}{18}(\frac{\partial \hat{n}}{\partial \delta})^2}{|\rho(s)|^3} \right\rangle_s \quad (17)$$

This can be written in the form

$$\frac{1}{\tau_{\text{dk}}} = \frac{1}{\tau_{\text{st}}} + \frac{1}{\tau_{\text{dep}}} \quad (18)$$

where  $\tau_{\text{st}}^{-1}$  can be (very well) approximated by  $\tau_{\text{bks}}^{-1}$  in (14) and

$$\tau_{\text{dep}}^{-1} = \frac{5\sqrt{3} r_e \gamma^5 \hbar}{8 m_e} \frac{1}{C} \oint ds \left\langle \frac{\frac{11}{18}(\frac{\partial \hat{n}}{\partial \delta})^2}{|\rho(s)|^3} \right\rangle_s \quad (19)$$

The time dependence for build-up from an initial polarization  $P_0$  to equilibrium is

$$P(t) = P_{\text{ens,dk}} \left[ 1 - e^{-t/\tau_{\text{dk}}} \right] + P_0 e^{-t/\tau_{\text{dk}}} \quad (20)$$

This formula can be used to calibrate polarimeters (Sec.7.6) [14]. However, the calibration will be imprecise if  $\frac{\partial \hat{n}}{\partial \delta}$  in the numerator of Eq.(15) is not well enough known. For examples of build-up curves see [5].

**Resonances** Away from the spin-orbit resonances<sup>2</sup> (see also Sec.2.6.7)

$$\nu_{\text{sp}} = k + k_x \nu_x + k_y \nu_y + k_z \nu_z \quad (21)$$

$\hat{n}(\vec{u}; s) \approx \hat{n}_0(s)$ . But near resonances  $\hat{n}(\vec{u}; s)$  deviates from  $\hat{n}_0(s)$  by typically tens of milliradians at a few tens of GeV and the deviation increases with distance in phase space from the closed orbit. The “*spin orbit coupling function*”  $\frac{\partial \hat{n}}{\partial \delta}$ , which quantifies the depolarization, can then be large and the equilibrium polarization can be low. Note that even very close to resonances,  $|\langle \hat{n} \rangle_s| \approx 1$ :  $\tilde{P}_{\text{ens,dk}}$  (Eq.(16)) is mainly influenced by the value of  $P_{\text{dk}}$  in Eq.(15).

To get high polarization, one must have  $(\partial \hat{n} / \partial \delta)^2 \ll 1$  in dipole magnets. The machine optimization required to make  $\frac{\partial \hat{n}}{\partial \delta}$  small is called “*spin matching*” (Sec.2.6.8).

**Kinetic polarization** The (numerator) term linear in  $\frac{\partial \hat{n}}{\partial \delta}$  in Eq.(15) is due to a correlation between the spin orientation and the radiation power [4]. In rings where  $\hat{n}_0$  is horizontal due, say, to the presence of a solenoid Siberian Snake (Sec.2.6.3) [15],  $\frac{\partial \hat{n}}{\partial \delta}$  has a vertical component in the dipole fields. This can lead to a build-up of polarization (“*kinetic polarization*”) even though the pure Sokolov-Ternov effect vanishes. The rate is  $\tau_{\text{dk}}^{-1}$ .

**Beam energy calibration** A polarized electron beam can be depolarized by applying a weak oscillating magnetic field perpendicular to  $\hat{n}_0$  with a frequency  $f_{\text{rf}}$  related to the fractional part of the spin tune  $\tilde{\nu}_{\text{sp}}$  by [19]

$$f_{\text{rf}} = f_0 \tilde{\nu}_{\text{sp}} \quad \text{or} \quad f_{\text{rf}} = f_0 (1 - \tilde{\nu}_{\text{sp}}) \quad (22)$$

Thus the required  $f_{\text{rf}}$  gives an accurate measurement of  $\tilde{\nu}_{\text{sp}}$  and this gives high relative precision on  $\nu_{\text{sp}}$ . By relating  $\nu_{\text{sp}}$  to the average energy of each beam, high precision measurements of the CM energy of colliding  $e^+e^-$  beams and of the

masses of the  $Z$  and of vector mesons such as the  $\Upsilon$ 's can be obtained [20, 21, 22, 23, 24]. High polarizations are not needed for these measurements.

**Notes** For a treatment of phase space and polarization evolution equations using an equation of the Fokker-Planck type (Sec.2.4.7), see [16, 17, 18]. For an overview of measurements see [25, 5, 26]. For more information and an overview of the theoretical background see [8, 18].

## References

- [1] A.A. Sokolov, I.M. Ternov, Sov. Phys. Dokl. 8 No.12 (1964) 1203
- [2] V.N. Baier, V.M. Katkov, Sov. Phys. JETP. 25 (1967) 944
- [3] V.N. Baier et al, Sov. Phys. JETP. 31 (1970) 908
- [4] B. Montague, Physics Reports 113 (1984)
- [5] D.P. Barber et al, Phys. Lett. 343B (1995) 436
- [6] V.N. Baier, Yu. Orlov, Sov. Phys. Dokl. 10 (1966) 1145
- [7] Ya.S. Derbenev, A.M. Kondratenko, Sov. Phys. JETP. 37 (1973) 968
- [8] D.P. Barber et al, ICFA workshop on Quantum Aspects of Beam Physics, World Scientific (1998)
- [9] D.P. Barber, J.A. Ellison, K. Heinemann, PRST-AB 7 (12), 124002 (2004)
- [10] G.H. Hoffstaetter, M. Vogt, D.P. Barber, PRST-AB 2 (11) 114001 (1999)
- [11] S.R. Mane, PR A36 (1987) 105
- [12] D.P. Barber, S.R. Mane, PR A37 (1988) 456
- [13] L.N. Hand, A. Skuja, PRL 59 (1987) 1910
- [14] D.P. Barber et al, NIM A338 (1994) 166
- [15] C.W. de Jager, V. Ptitsin, Yu.M. Shatunov, Proc. 12th Int. Symp. High Energy Spin Physics, World Scientific (1997)
- [16] K. Heinemann, DESY Report 97-166 (1997) and arXiv: physics/9709025
- [17] K. Heinemann, D.P. Barber, NIM A463 (2001) 62 and A469 (2001) 294
- [18] D.P. Barber, G. Ripken, arXiv: physics/9907034v2
- [19] Kh.A. Simonian, Yu.F. Orlov, Sov. Phys. JETP. 18 (1964) 123
- [20] Yu.M. Shatunov, A.N. Skrinsky, Particle World 1 (1989) 35
- [21] R. Assmann et al, Z. f. Physik C66 (1995) 567
- [22] M. Placidi et al, EPAC98
- [23] D.P. Barber et al, PL B135 (1984) 498
- [24] W.W. MacKay et al, PR D29 (1984) 2483
- [25] D.P. Barber, Proc. 12th Int. Symp. High Energy Spin Physics, loc.cit.
- [26] Yu.M. Shatunov, PA 32 (1990) 139

<sup>2</sup>The resonance condition should be more precisely expressed in terms of the “*amplitude dependent spin tune*” [8, 9]. But for typical  $e^-/e^+$  rings the amplitude dependent spin tune differs only insignificantly from  $\nu_{\text{sp}}$ .

## 2.6.7 Computing Algorithms for $e^-/e^+$ Polarization in Storage Rings

D.P. Barber, DESY  
G. Ripken, Deceased 2004

There are two classes of computer algorithm for estimating the equilibrium  $e^-/e^+$  polarization in real rings (for some notations see Sec.2.6.6):

(i) Methods based on evaluating  $\frac{\partial \hat{n}}{\partial \delta}$  in the Derbenev-Kondratenko-Mane (DKM) formula (Sec.2.6.6) given the ring layout and magnet strengths; and

(ii) algorithms such as SLICKTRACK [1] estimate  $\tau_{\text{dep}}$  using Monte-Carlo tracking.

The class (i) algorithms are further divided according to the degree of linearization of the spin and orbital motion:

(ia) The SLIM family (SLIM [2, 3] - thin lenses, SLICK [4] - thick lenses), and ASPIRRIN [5]: linearized orbital and spin motion;

(ib) SMILE [6] and SODOM [7]: Linearized orbital motion but  $|\vec{\xi}|$ -preserving spin motion;

(ic) SpinLie: Nonlinear orbital motion and  $|\vec{\xi}|$ -preserving spin motion (Sec.2.6.9).

**The linear approximation – SLIM** The orbital and spin motion are linearized and the components  $\omega_z^{\text{sb}}$ ,  $\omega_x^{\text{sb}}$  and  $\omega_y^{\text{sb}}$  of  $\vec{\omega}^{\text{sb}}$  are linearized in the form [8, 9]

$$\begin{bmatrix} \omega_z^{\text{sb}} \\ \omega_x^{\text{sb}} \\ \omega_y^{\text{sb}} \end{bmatrix} = \mathbf{F}_{3 \times 6} \begin{bmatrix} x \\ p_x \\ y \\ p_y \\ z \\ \delta \end{bmatrix} \quad (1)$$

where  $\vec{u} \equiv (x, p_x, y, p_y, z, \delta)$  describes motion with respect to the closed orbit. In particular  $p_x = x'$  and  $p_y = y'$  (except in solenoids).

The detailed forms of the matrix  $\mathbf{F}_{3 \times 6}$  for the various machine elements can be found in [9]. The orbital motion in sextupoles is linearized. For example for a quadrupole, defining  $\tilde{g} = -(1 + a\gamma_0)g$  where  $g = \frac{e}{p_0} \frac{\partial B_y}{\partial x}$

$$\mathbf{F}(s) = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \tilde{g} & 0 & 0 & 0 \\ \tilde{g} & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad (2)$$

In linear approximation we write

$$\hat{n}(\vec{u}; s) = \hat{n}_0(s) + \alpha(\vec{u}; s)\hat{m}(s) + \beta(\vec{u}; s)\hat{l}(s) \quad (3)$$

valid for  $\sqrt{\alpha^2 + \beta^2} \ll 1$ .

The combined orbit and spin motion is then described by  $8 \times 8$  transport matrices of the form

$$\hat{\mathbf{M}} = \begin{bmatrix} \mathbf{M}_{6 \times 6} & \mathbf{0}_{6 \times 2} \\ \mathbf{G}_{2 \times 6} & \mathbf{D}_{2 \times 2} \end{bmatrix} \quad (4)$$

acting on the vector  $(\vec{u}, \alpha, \beta)$ , where  $\mathbf{M}_{6 \times 6}$  is a symplectic matrix describing orbital motion and  $\mathbf{G}_{2 \times 6}$  describes the coupling of the spin variables  $(\alpha, \beta)$  to the trajectory and depends on  $\hat{m}(s)$  and  $\hat{l}(s)$  (Eq.(3), Sec.2.6.8).  $\mathbf{D}_{2 \times 2}$  is a rotation matrix associated with the spin basis rotation of Eq.(11) in Sec.2.6.6 [8, 9].

The eigenvectors for one turn defined by  $\hat{\mathbf{M}}(s_0 + C, s_0) \cdot \vec{q}_\mu = \hat{\lambda}_\mu \cdot \vec{q}_\mu$  are written in the form

$$\begin{aligned} \vec{q}_k(s_0) &= \begin{bmatrix} \vec{v}_k(s_0) \\ \vec{w}_k(s_0) \end{bmatrix}, \quad \vec{q}_{-k}(s_0) = [\vec{q}_k(s_0)]^* \\ &\quad \text{for } k = I, II, III, \\ \vec{q}_k(s_0) &= \begin{bmatrix} \vec{0}_6(s_0) \\ \vec{w}_k(s_0) \end{bmatrix}, \quad \vec{q}_{-k}(s_0) = [\vec{q}_k(s_0)]^* \\ &\quad \text{for } k = IV \end{aligned} \quad (5)$$

for arbitrary  $s_0$ . The  $\vec{v}_k$  are the eigenvectors for orbital motion with eigenvalues  $\lambda_k = e^{-i2\pi\nu_k}$  and with  $\nu_{-k} = -\nu_k$  ( $k = I, II, III$ ). These eigenvectors obey the orthogonality and normalization relations [2, 8]. The corresponding eigenvalues of  $\hat{\mathbf{M}}(s_0 + C, s_0)$  are  $\hat{\lambda}_k = \lambda_k$  ( $k = I, II, III$ ) and  $\hat{\lambda}_{IV} = e^{-i2\pi\nu_{IV}}$  with  $\nu_{IV} = \nu_{\text{sp}}$  and with  $\nu_{-IV} = -\nu_{IV}$ .

The spin parts of the eigenvectors  $\vec{w}_k(s_0)$  ( $k = I, II, III$ ) and  $\vec{w}_{IV}(s_0)$  can be written

$$\begin{aligned} \vec{w}_k(s_0) &= -[\mathbf{D}(s_0 + C, s_0) - \hat{\lambda}_k]^{-1} \\ &\quad \times \mathbf{G}(s_0 + C, s_0) \vec{v}_k(s_0) \\ &\quad \text{for } k = I, II, III \end{aligned} \quad (6)$$

$$\vec{w}_{IV}(s_0) = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -i \end{bmatrix} e^{-i\psi_{\text{sp}}(s_0)} \quad \text{for } k = IV$$

and  $\vec{w}_{-k}(s_0) = [\vec{w}_k(s_0)]^*$ , ( $k = I, II, III, IV$ ).

In this linear approximation  $\hat{n}(\vec{u}; s)$  can be obtained via [10, 11, 9]

$$\begin{aligned} \hat{n}(\vec{u}; s) - \hat{n}_0(s) &\equiv \begin{bmatrix} \alpha(\vec{u}; s) \\ \beta(\vec{u}; s) \end{bmatrix} \\ &= \sum_{k=I,II,III} \{A_k \vec{w}_k(s) + A_{-k} \vec{w}_{-k}(s)\} \end{aligned} \quad (7)$$

for amplitudes  $A_k$  determined by the orbit via

$$\vec{u}(s) = \sum_{k=I,II,III} \{A_k \vec{v}_k(s) + A_{-k} \vec{v}_{-k}(s)\} \quad (8)$$

Then with respect to the  $(\hat{n}_0, \hat{m}, \hat{l})$  frame,

$$\begin{aligned} \frac{\partial \hat{n}}{\partial \delta} &\equiv i \sum_{k=I,II,III} \{v_{k5}^* \vec{w}_k - v_{k5} \vec{w}_k^*\} \\ &= -2 \operatorname{Im} \sum_{k=I,II,III} v_{k5}^* \vec{w}_k \end{aligned} \quad (9)$$

The depolarization time is then

$$\begin{aligned} \tau_{\text{dep,lin}}^{-1} &\approx \frac{55\sqrt{3}}{36} \frac{r_e \gamma_0^5 \hbar}{m_e C} \int_{s_0}^{s_0+C} d\tilde{s} \frac{1}{|\rho(\tilde{s})|^3} \\ &\times \sum_{\mu=1}^2 \left( \operatorname{Im} \sum_{k=I-III} [v_{k5}^*(\tilde{s}) w_{k\mu}(\tilde{s})] \right)^2 \end{aligned} \quad (10)$$

In this linearized theory the vectors  $\hat{n}(\vec{u}; s)$  and  $\frac{\partial \hat{n}}{\partial \delta}$  display only the first order resonances

$$\nu_{\text{sp}} = k + k_I \nu_I + k_{II} \nu_{II} + k_{III} \nu_{III} \quad (11)$$

with  $|k_I| + |k_{II}| + |k_{III}| = 1$ . They arise from the denominator matrix in Eq.(6). The formalism is only valid for  $\sqrt{\alpha^2 + \beta^2} \ll 1$ . In a fully uncoupled optic  $I \rightarrow x$ ,  $II \rightarrow y$  and  $III \rightarrow z$ . This formalism forms the natural language for maximizing the polarization by “*spin matching*” (Sec.2.6.8).

**Other computer codes [12]** The SMILE algorithm works in the thin lens approximation and calculates  $\frac{\partial \hat{n}}{\partial \delta}$  by extending the first-order perturbation theory of SLIM to high-order for the spin motion. High-order resonance effects are manifested by high-order resonance denominators. SODOM represents  $\hat{n}$  in terms of spinors and encodes the condition  $\hat{n}(\vec{u}; s) = \hat{n}(\vec{u}; s + C)$  and the similar periodicity of the 1-turn  $2 \times 2$  spinor transfer matrix, in terms of  $2\pi$ -periodicity in the three orbital phases.  $\hat{n}(\vec{u}; s)$  is then constructed by solving for the Fourier coefficients and  $\frac{\partial \hat{n}}{\partial \delta}$  is obtained by numerical differentiation. The algorithm SpinLie utilizes Lie-algebraic methods to provide a perturbation expansion for  $\hat{n}$  and can handle moderately nonlinear orbit motion. The vector  $\hat{n}(\vec{u}; s)$  can also be obtained by “*stroboscopic averaging*” using the code SPRINT [12]. This algorithm automatically includes all orders of resonance.

The above analytical algorithms all exploit the DKM formula (Sec.2.6.6) but SLICKTRACK (an extended version of SLICK) uses the transfer matrices of SLICK to simulate the depolarization process directly using Monte-Carlo tracking simulations of the effects on the trajectories, and then on the spins, of stochastic photon emission and

damping, and delivers estimates of  $\tau_{\text{dep}}$ . The polarization is then obtained from the approximation (Sec.2.6.6)

$$P_{\text{eq}} = P_{\text{bks}} \frac{\tau_{\text{tot}}}{\tau_{\text{bks}}} \text{ with } \frac{1}{\tau_{\text{tot}}} = \frac{1}{\tau_{\text{bks}}} + \frac{1}{\tau_{\text{dep}}} \quad (12)$$

This ignores the (normally small) term with  $\frac{\partial \hat{n}}{\partial \delta}$  in the numerator of the DKM formula. SLICK-TRACK calculates with full  $|\vec{\xi}|$ -preserving spin motion and, in contrast to the analytical algorithms, it has the advantage that it can handle the strongly nonlinear orbital motion caused by beam-beam forces.

## References

- [1] D.P. Barber, Proc. 16th Int. Spin Physics Symp., World Scientific (2005) 738
  - [2] A.W. Chao, NIM 180 (1981) 29(\*)
  - [3] A.W. Chao, AIP Proc. 87 (1981) 395(\*)
  - [4] SLICK is a thick lens version of SLIM by D.P. Barber (1982)
  - [5] C.W. de Jager, V. Ptitsin, Yu.M. Shatunov, Proc. 12th Int. Symp. High Energy Spin Physics, World Scientific (1997)
  - [6] S.R. Mane, PRA 36 (1987) 105
  - [7] K. Yokoya, KEK Report 92-6 (1992), DESY Report 99-006 (1999)
  - [8] H. Mais, G. Ripken, DESY Report 83-62 (1983)(\*)
  - [9] D.P. Barber, K. Heinemann, G. Ripken, Z. f. Physik C64 (1994) 117
  - [10] S.R. Mane, DESY Report 85-125 (1985); Cornell Report CLNS-86/750
  - [11] D.P. Barber, K. Heinemann, G. Ripken, DESY Report M-92-04 (1992), revision 1999
  - [12] G.H. Hoffstaetter, M. Vogt, D.P. Barber, PRST-AB 2 (11) 114001 (1999)
- (\*) Modern notation: replace  $\hat{n}$  by  $\hat{n}_0$

### 2.6.8 Spin Matching in $e^-/e^+$ Rings

D.P. Barber, DESY  
G. Ripken, Deceased 2004

**Spin matching in the SLIM formalism** Synchrotron radiation causes depolarization and to maximize the polarization we must minimize  $\tau_{\text{dep}}^{-1}$  (See Sec.2.6.6 for some notations). In lowest approximation this is quantified by the SLIM formalism (Sec.2.6.7). The manipulations are called spin matching and in real rings this takes place in stages as follows.

Stage 1: *Strong synchrobeta spin matching of the perfectly aligned ring*

## Sec.2.6: POLARIZATION

To minimize  $\tau_{\text{dep}}^{-1}$  we must minimize  $v_{k5}^*$  ( $k = I, II, III$ ) or the components of  $\vec{w}_k$  at azimuths where  $1/|\rho(s)|^3$  is large. The  $v_{k5}^*$  determine the orbit excitation due to synchrotron radiation [1]. In particular, for rings without  $x$ - $y$  coupling,  $v_{II5}^*$  usually vanishes in the arcs since the vertical dispersion  $\eta_y$  vanishes. However,  $v_{II5}^*$  does not vanish inside spin rotators (Sec.2.6.3) containing vertical bends. On the other hand  $v_{I5}^*$  tends not to vanish in the arcs since the horizontal dispersion  $\eta_x \neq 0$ . Finally,  $v_{III5}^*$  essentially never vanishes. Each case must be evaluated individually but the minimal recipe is to try to minimize  $\vec{w}_k$  for ( $k = I, II, III$ ) only at azimuths where  $|v_{k5}(s)|^2/|\rho(s)|^3$  is sufficiently large. This in turn requires (Eq.(6), Sec.2.6.7) that  $\mathbf{G}(s + C, s) \cdot \vec{v}_k(s)$  be minimized for  $k = I, II, III$ . The ring layout must be designed with this in mind and sufficient independently powered quadrupoles be provided to ensure flexibility in the optics. Subsequent calculations with SLIM will indicate whether the match criteria for the adopted design suffice. Consider, for example, a specific mode  $k$ . Label those bending magnets at which  $|v_{k5}(s)|^2/|\rho(s)|^3$  is large by  $\mu_1^{(k)}, \mu_2^{(k)}, \dots, \mu_{n_k}^{(k)}$ . Then the suppression of depolarization associated with the  $k$ th mode requires that  $\vec{w}_k(s_{\mu_i}) = 0$  for all ( $i = 1$  to  $n_k$ ). In general this in turn requires [2]

$$\begin{aligned}\mathbf{G}(s_{\mu_2}, s_{\mu_1}) \vec{v}_k(s_{\mu_1}) &= 0 \\ \mathbf{G}(s_{\mu_3}, s_{\mu_2}) \vec{v}_k(s_{\mu_2}) &= 0 \\ &\vdots \\ \mathbf{G}(s_{\mu_1} + C, s_{\mu_n}) \vec{v}_k(s_{\mu_n}) &= 0\end{aligned}\quad (1)$$

where we suppressed the superscript label “ $k$ ”. To fulfill Eq.(1) we then require the  $G_{ij}(s_{\mu_{l+1}}, s_{\mu_l})$  to vanish when the  $j$ th component of  $\vec{v}_k$  does not vanish. The matrix  $\mathbf{G}$  can be written in the form

$$\mathbf{G}(s_2, s_1) = \int_{s_1}^{s_2} d\tilde{s} \, \mathbf{D}(s_2, \tilde{s}) \mathbf{G}_0(\tilde{s}) \mathbf{M}(\tilde{s}, s_1)$$

where

$$\mathbf{G}_0 = \begin{bmatrix} l_s & l_x & l_y \\ -m_s & -m_x & -m_y \end{bmatrix} \cdot \mathbf{F} \quad (2)$$

Thus  $G_{ij}(s_{\mu_{l+1}}, s_{\mu_l})$  depends on the orientation of  $(\hat{m}, \hat{l})$  so that in some cases some elements of  $G_{ij}(s_{\mu_{l+1}}, s_{\mu_l})$  vanish automatically. But in general these conditions can only be fulfilled by adjusting quadrupole strengths – while maintaining other necessary features of the orbital optics.

This is *strong synchrobeta spin matching*. A section of the ring satisfying a condition in Eq.(1) is *spin transparent* for mode  $k$ . The interpretation is immediate: the overall spin-orbit coupling for the section vanishes for mode  $k$ . Clearly, the exact spin matching conditions depend on the layout of the ring and each case must be handled individually. In thin lens approximation the  $\mathbf{G}$  matrix for a quadrupole of length  $l_q$  is

$$\mathbf{G} = \begin{bmatrix} -\tilde{q}l_y & 0 & -\tilde{q}l_x & 0 & 0 & 0 \\ +\tilde{q}m_y & 0 & +\tilde{q}m_x & 0 & 0 & 0 \end{bmatrix} \quad (3)$$

where  $\tilde{q} = (1 + \alpha\gamma) g l_q$ . The thin- and thick-lens forms of  $\mathbf{G}$  for other magnet types are given in [3, 4, 5].

If the  $G_{ij}(s_{\mu_{l+1}}, s_{\mu_l})$  cannot be brought to zero while maintaining an acceptable optic, then the  $\mathbf{G}(s_{\mu_l} + C, s_{\mu_l}) \vec{v}_k(s_{\mu_l})$  themselves should be minimized. This essentially means that the effects of elements of the  $\mathbf{G}$  matrices of sections of the ring are made to partially cancel one another. Spin matching of a ring with a solenoid Siberian Snake (Sec.2.6.3) is an example of this [6]. By Eq.(6), Sec.2.6.7, reduction of  $\mathbf{G}(s + C, s) \vec{v}_k(s)$  for ( $k = I, II, III$ ) also reduces the angle between  $\hat{n}$  and  $\hat{n}_0$  at azimuth  $s$ .

*Alternative Stage 1: Harmonic synchrobeta spin matching of the perfectly aligned ring*

If the strong spin matching methods are impractical for some reason, another approach aimed at minimizing the strengths of depolarizing resonances can be adopted [7, 8].

*Reformulation in terms of beta functions and dispersion [9]*

We can reformulate Stage 1 by transforming the coordinates from  $\vec{u} \equiv (x, p_x, y, p_y, z, \delta)$  to  $\vec{\tilde{u}} \equiv (\tilde{x}, \tilde{p}_x, \tilde{y}, \tilde{p}_y, \tilde{z}, \delta)$  defined by

$$\vec{\tilde{u}} = \mathbf{K} \vec{u}$$

with the symplectic matrix

$$\mathbf{K}(s) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & -\eta_1 \\ 0 & 1 & 0 & 0 & 0 & -\eta_2 \\ 0 & 0 & 1 & 0 & 0 & -\eta_3 \\ 0 & 0 & 0 & 1 & 0 & -\eta_4 \\ \eta_2 & -\eta_1 & \eta_4 & -\eta_3 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

where the  $\vec{\eta} \equiv (\eta_1, \eta_2, \eta_3, \eta_4) = (\eta_x, \eta'_x, \eta_y, \eta'_y)$  are components of the dispersion vector. Then

$$\tilde{x} = x - \delta\eta_x, \quad \tilde{y} = y - \delta\eta_y$$

The formalism remains canonical and in particular, there are new eigenvectors

$$\implies \vec{v}_{\mu}(s) = \mathbf{K}(s) \vec{v}_{\mu}(s)$$

for which the eigenvalues and orthogonality conditions are unchanged. Furthermore  $\tilde{\mathbf{G}}(s_2, s_1) = \mathbf{G}(s_2, s_1)\mathbf{K}^{-1}(s_1)$ .

The depolarization rate then takes the form

$$\tau_{\text{dep,lin}}^{-1} = \frac{55\sqrt{3}}{36} \frac{r_e \gamma^5 \hbar}{m_e} \frac{1}{C} \int_{s_0}^{s_0+C} d\tilde{s} \frac{1}{|\rho(\tilde{s})|^3} \\ \times \sum_{\mu=1}^2 \left( \text{Im} \sum_{k=I-III} [f_k(\tilde{s}) \tilde{w}_{k\mu}(\tilde{s})] \right)^2$$

with  $f_k = \sum_{n=1}^6 [K^{-1}]_{5n} \tilde{v}_{kn}^* = v_{k5}^*$  and  $\tilde{w}_k = \vec{w}_k$ . In the special case, or the approximation, of no orbital coupling, the  $6 \times 6$  orbit matrices just consist of three  $2 \times 2$  matrices on the diagonal. The eigenvectors  $\vec{v}_k(s)$  of the revolution matrix are then written in the form

$$\vec{v}_I = \begin{bmatrix} \vec{t}_x \\ \vec{0}_2 \\ \vec{0}_2 \end{bmatrix}, \vec{v}_{II} = \begin{bmatrix} \vec{0}_2 \\ \vec{t}_y \\ \vec{0}_2 \end{bmatrix}, \vec{v}_{III} = \begin{bmatrix} \vec{0}_2 \\ \vec{0}_2 \\ \vec{t}_z \end{bmatrix} \\ \vec{t}_r = \frac{1}{\sqrt{2\beta_r(s)}} \begin{bmatrix} \beta_r(s) \\ -[\alpha_r(s) + i] \end{bmatrix} e^{-i\psi_r(s)}$$

( $r \equiv x, y, z$ ) and the  $f_k$  are given by  $f_I \equiv f_x = -(\tilde{v}_{I1}\eta_2 - \tilde{v}_{I2}\eta_1)$ ;  $f_{II} \equiv f_y = -(\tilde{v}_{II3}\eta_4 - \tilde{v}_{II4}\eta_3)$  and  $f_{III}(s) \equiv f_z = \sqrt{\frac{\beta_z}{2}} e^{-i\psi_z(s)}$ . The  $|f_I|^2$  and  $|f_{II}|^2$  are just the factors

$$\frac{\eta_r^2 + (\alpha_r\eta_r + \beta_r\eta_r)^2}{\beta_r} \quad (r = x, y)$$

in [10].  $f_{III}$  is in practice almost independent of  $s$ . Note that  $\alpha$ 's and  $\beta$ 's here are Courant-Snyder parameters, not the spin components. With these coordinates we can write

$$\vec{w}_k(s_0) = - \left[ \mathbf{D}(s_0 + C, s_0) - \hat{\lambda}_k \right]^{-1} \\ \times \tilde{\mathbf{G}}(s_0 + C, s_0) \vec{v}_k(s_0)$$

for ( $k = I, II, III$ ).

In thin lens approximation the  $\tilde{\mathbf{G}}$  matrix for a quadrupole is

$$\tilde{\mathbf{G}} = \begin{bmatrix} -\tilde{q}l_y & 0 & -\tilde{q}l_x & 0 & 0 & \kappa_1 \\ +\tilde{q}m_y & 0 & +\tilde{q}m_x & 0 & 0 & \kappa_2 \end{bmatrix}$$

where  $\kappa_1 = -\tilde{q}l_y\eta_1 - \tilde{q}l_x\eta_3$  and  $\kappa_2 = +\tilde{q}m_y\eta_1 + \tilde{q}m_x\eta_3$ . By separating the transverse coordinates into betatron and dispersion contributions, column six of  $\tilde{\mathbf{G}}$  contains terms depending on dispersions.

The strong spin matching condition  $\tilde{w}_k = 0$  now means setting the  $\tilde{\mathbf{G}}(s_{\mu_{l+1}}, s_{\mu_l}) \vec{v}_k(s_{\mu_l}) = 0$  in analogy with Eq.(1). Then in a fully uncoupled optic and by taking into account only the quadrupoles this is equivalent to requiring [11, 12]:

For horizontal motion:

$$-\frac{(1+a\gamma)}{\sqrt{2}} \frac{1}{C} \int_{s_{\mu_l}}^{s_{\mu_{l+1}}} d\tilde{s} \sqrt{\beta_x(\tilde{s})} g(\tilde{s}) e^{-i\psi_x(\tilde{s})} \\ \times [l_y(\tilde{s}) \pm i m_y(\tilde{s})] e^{\mp i\psi_{\text{sp}}(\tilde{s})} = 0 \quad (4)$$

For vertical motion:

$$-\frac{(1+a\gamma)}{\sqrt{2}} \frac{1}{C} \int_{s_{\mu_l}}^{s_{\mu_{l+1}}} d\tilde{s} \sqrt{\beta_y(\tilde{s})} g(\tilde{s}) e^{-i\psi_y(\tilde{s})} \\ \times [l_x(\tilde{s}) \pm i m_x(\tilde{s})] e^{\mp i\psi_{\text{sp}}(\tilde{s})} = 0 \quad (5)$$

For longitudinal motion:

$$-\frac{(1+a\gamma)}{\sqrt{2}} \frac{1}{C} \int_{s_{\mu_l}}^{s_{\mu_{l+1}}} d\tilde{s} \frac{[\alpha_z(\tilde{s}) + i]}{\sqrt{\beta_z(\tilde{s})}} \\ \times g(\tilde{s}) e^{-i\psi_z(\tilde{s})} [\eta_y [l_x(\tilde{s}) \pm i m_x(\tilde{s})]] \\ + \eta_x [l_y(\tilde{s}) \pm i m_y(\tilde{s})] e^{\mp i\psi_{\text{sp}}(\tilde{s})} = 0 \quad (6)$$

Harmonic synchrobeta spin matching in terms of beta functions and dispersion follows the path detailed earlier under “Alternative Stage 1” but with the eigenvectors  $\vec{v}_k$  and the matrices  $\tilde{\mathbf{F}}$ . Typical expressions can be found in [12, 13, 7].

#### Commentary

A detailed commentary on the practical use of this formalism, with examples, can be found in [7, 8].

#### Stage 2: Harmonic closed orbit spin matching

Once the perfectly aligned ring has been spin matched, the effects of misalignment must be addressed. In a perfectly aligned flat ring with no solenoids,  $\tau_{\text{dep,lin}}^{-1}$  is automatically zero [7, 8]. In real misaligned rings there is a vertical closed orbit distortion and  $\hat{n}_0$  is tilted from the vertical in the arcs (see below). Then parts of  $\mathbf{G}$  and  $\tilde{\mathbf{G}}$  for the arc quadrupoles do not vanish so that normally  $\tau_{\text{dep,lin}}^{-1} \neq 0$ . In practice the tilts can be tens of milliradians but even these small angles can lead to strong depolarization so that it is essential that the ring be very well aligned from the beginning. Note that *vertical* closed orbit distortion leads primarily to depolarization due to *horizontal* synchrobetatron motion in the arcs. Note also that tilts of tens of milliradians cause a negligible decrease of the underlying ST polarization (Eq.(13), Sec.2.6.6).

The key to minimizing depolarization due to distortions is reduction of the tilt of  $\hat{n}_0$  in the arcs.

This is conveniently done using SLIM-like concepts [14]. Viewed from the  $(\hat{n}_0, \hat{m}, \hat{l})$  frame calculated for the design orbit, the first order deviation of  $\hat{n}_0$  from the design orientation can be written as

$$[\delta n_{01}(s) - i\delta n_{02}(s)] = -i\frac{C}{2\pi} \sum_k h_k \frac{e^{i2\pi ks/C}}{k - \nu_{sp}}$$

where the  $h_k$  are Fourier coefficients given by

$$h_k = \frac{1}{C} \int_{s_0}^{s_0+C} d\tilde{s} [d_1(\tilde{s}) - id_2(\tilde{s})] e^{-ik2\pi\tilde{s}/C}$$

Here

$$\begin{bmatrix} d_1 \\ d_2 \end{bmatrix} = \begin{bmatrix} l_s & l_x & l_y \\ -m_s & -m_x & -m_y \end{bmatrix} \cdot \left\{ \mathbf{F} \cdot \vec{u}_{co} - \frac{e}{p_0} \begin{bmatrix} \Delta B_s \frac{1+\alpha\gamma_0}{1+\gamma_0} \\ \Delta B_x(1+\alpha\gamma_0) \\ \Delta B_y(1+\alpha\gamma_0) \end{bmatrix} \right\}$$

where the  $\Delta B_{x,y,s}$  are field errors and  $\vec{u}_{co}$  is the deviation of the 6-D closed orbit from the design orbit.  $\delta\hat{n}_0$  can be minimized by using correction coils to adjust the closed orbit (e.g. by generating closed bumps so that the luminosity is not affected) in such a way that the real and imaginary parts of  $h_k$ , with  $k$  near  $\nu_{sp}$ , are small. This is *harmonic closed orbit spin matching*. Since the ring misalignments are usually not well known and the closed orbit cannot usually be measured well enough, the closed orbit correction is carried out empirically by observing the polarization [15]. If the misalignments and the orbit are well enough known the correction coil strength can be calculated *ab initio* (deterministic harmonic closed orbit spin matching) [16]. The correction scheme should be chosen so that it achieves the maximum effect on  $\delta\hat{n}_0$  with the smallest possible additional orbit distortion.

#### *Higher order resonances and Beam-beam spin matching*

By calculating beyond the linearizations of the SLIM formalism one finds higher order spin-orbit resonances. The beam-beam interaction is equivalent to a nonlinear lens and can spoil a spin match. For comments and more details see [7, 8].

## References

- [1] A.W. Chao, J. Appl. Phys. 50 (1979) 595
  - [2] H. Mais, G. Ripken, DESY Report M-84-04 (1984)<sup>(\*)</sup>
  - [3] A.W. Chao, NIM 180 (1981) 29<sup>(\*)</sup>
  - [4] A.W. Chao, AIP Proc. 87 (1981) 395<sup>(\*)</sup>
  - [5] D.P. Barber et al, PA 17 (1985) 243<sup>(\*)</sup>
  - [6] C.W. de Jager, V. Ptitsin, Yu.M. Shatunov, Proc. 12th Int. Symp. High Energy Spin Physics, World Scientific (1997)
  - [7] D.P. Barber, G. Ripken, arXiv: physics/9907034v2
  - [8] Handbook of Accelerator Physics and Engineering, 1st ed., World Scientific, 2006
  - [9] H. Mais, G. Ripken, DESY Report 86-29 (1986)<sup>(\*)</sup>
  - [10] M. Sands, SLAC 121 (1970)
  - [11] A.W. Chao, K. Yokoya, KEK Report 81-7 (1981)
  - [12] J. Buon, K. Steffen, NIM A245 (1986) 248
  - [13] K. Yokoya, KEK Report 81-19 (1982)
  - [14] D.P. Barber et al, DESY Report 85-44 (1985)<sup>(\*)</sup>
  - [15] D.P. Barber et al, NIM A338 (1994) 166
  - [16] R. Assmann et al, Proc. 11th Int. Symp. High Energy Spin Physics, AIP Proc. 343 (1995)
- (\*) Modern notation: replace  $\hat{n}$  by  $\hat{n}_0$ .

### 2.6.9 Lie Algebra for Spin Motion

K. Yokoya, KEK

The Lie algebra method (Sec.2.3.6) can include the classical spin motion [1]. The Hamiltonian is written in general as

$$\mathcal{H} = \mathcal{H}_{orb}(z, \theta) + \mathbf{s} \cdot \mathcal{W}(z, \theta)$$

where  $z$  represents the six orbital variables,  $\theta$  is the machine azimuth,  $\mathcal{H}_{orb}$  the orbit Hamiltonian,  $\mathbf{s}$  the spin vector, and  $\mathcal{W}$  is written in terms of the electromagnetic field along the orbit. The Thomas-BMT equation (Sec.2.6.1) is then

$$\frac{ds}{d\theta} = [\mathbf{s}, \mathcal{H}] = \mathcal{W} \times \mathbf{s}$$

where  $[\cdot]$  is the Poisson bracket. Note  $[s_i, s_j] = \sum_k \epsilon_{ijk} s_k$ . The Lie map of an element of length  $\Delta\theta$  is  $e^{-\mathcal{H}\Delta\theta}$ : (valid if  $\mathcal{H}$  is independent of  $\theta$  or if  $\Delta\theta$  is infinitesimally small). If one ignores the Stern-Gerlach force (the effect of spin on the orbit motion), the map can be decomposed as  $e^{-s \cdot \mathbf{W}} e^{-H}$ : where  $H$  and  $\mathbf{W}$  are polynomials of  $z$ . One can concatenate the Lie maps to obtain the map of a beamline. In contrast to the orbit case the concatenation of the spin map does not require the Baker-Campbell-Hausdorff formula; it can be written explicitly as

$$e^{-s \cdot \mathbf{W}_1} e^{-s \cdot \mathbf{W}_2} = e^{-s \cdot \mathbf{W}}$$

$$\cos \phi = \cos \phi_1 \cos \phi_2 - \mathbf{a}_1 \cdot \mathbf{a}$$

$$\frac{\mathbf{W}}{|\mathbf{W}|} \sin \phi = \mathbf{a}_2 \cos \phi_1 + \mathbf{a}_1 \cos \phi_2 + \mathbf{a}_1 \times \mathbf{a}_2$$

where  $\phi_1 = |\mathbf{W}_1|/2$ ,  $a_1 = (\mathbf{W}_1/|\mathbf{W}_1|) \sin \phi_1$ , etc. When the orbit one-turn map can be transformed to the normal form  $e^{-:V:}e^{-:H:}e^{:V:}$  where  $H$  is a function of the orbit action variables  $(z_{2i}^2 + z_{2i-1}^2)/2$  ( $i = 1, 2, 3$ ) and  $V = V(z, \theta)$ , the total map can also be written in normal form with the same  $H$  and  $V$ ,

$$e^{-:V:}e^{-:s\cdot U:}e^{-:H+s\cdot W:}e^{:s\cdot U:}e^{:V:}$$

where  $\mathbf{W}$  is a function of orbit action and is identified as the spin tune (times  $2\pi$ ). The Derbenev-Kondratenko vector  $\mathbf{n}(z, \theta)$  is given by  $s \cdot \mathbf{n} = e^{-:V:}e^{-:s\cdot U:}s_3$ . A computer code to evaluate the equilibrium polarization in electron storage rings using Lie algebra is available [2].

## References

- [1] K. Yokoya, NIM A258(1987)149
- [2] Yu. Eidelman, V. Yakimemko, PA 45(1994)17; 50(1995)261

## 2.7 BEAM COOLING

### 2.7.1 Stochastic Cooling

*M. Blaskiewicz, BNL*

Stochastic Cooling was invented by Simon van der Meer [1] and was demonstrated at the CERN ISR and ICE (Initial Cooling Experiment) [2]. Operational systems were developed at Fermilab [3] and CERN [4]. A complete theory of cooling of unbunched beams was developed [4, 5, 6], and was applied at CERN and Fermilab. Several new and existing rings employ coasting beam cooling [7].

Bunched beam cooling was demonstrated in ICE and has been observed in several rings designed for coasting beam cooling. High energy bunched beams have proven more difficult. Signal suppression was achieved in the Tevatron [8], though operational cooling was not pursued at Fermilab. Longitudinal cooling was achieved in the RHIC collider [9]. More recently a vertical cooling system in RHIC cooled both transverse dimensions via betatron coupling [10].

#### 2.7.1.1 Cooling rates

**Unbunched beam** A simple, but useful, estimate of stochastic cooling rates is

$$\frac{1}{\epsilon} \frac{d\epsilon}{dt} = -\frac{W}{N}[2g - g^2(M + U)] \quad (1)$$

where  $\epsilon$  is the beam emittance,  $N$  is the number of particles in the ring,  $W = f_{\max} - f_{\min}$  with  $f_{\max,\min}$  the frequency limits of the bandwidth  $W$ . The fastest cooling is obtained at the optimum system gain, i.e.,  $g = 1/(M + U)$ . The mixing factor  $M$  may be defined to be the ratio of the peak Schottky power density to the average Schottky power density (averaged over all the Schottky bands in the cooling system bandwidth).  $U$  is the ratio of electronic noise power to the average Schottky power density.

**Transverse cooling rate** Applying kinetic theory and feedback theory, the betatron emittance cooling rate for particles with revolution frequency  $f = \omega/2\pi$  is [5]

$$S_\omega = \sum_{m, \pm Q} \left\{ \frac{G[(m \pm Q)\omega]}{\epsilon_T[(m \pm Q)\omega]} \frac{e^{im(\theta_p - \theta_k) \pm i\psi_\beta}}{\pm i} + \frac{\pi N f(\omega)}{|m \pm Q|} \left| \frac{G[(m \pm Q)\omega]}{\epsilon_T[(m \pm Q)\omega]} \right|^2 \right\} \quad (2)$$

with  $Q$  the betatron tune,  $f(\omega)$  the normalized frequency distribution,  $G$  the electronic gain, and  $\psi_\beta$  the betatron phase between pickup and kicker. The dielectric response is

$$\epsilon_T[(m \pm Q)\omega] = 1 + NG((m \pm Q)\omega) \times \sum_{\ell, \pm \nu} \int d\omega_1 \frac{f(\omega_1) \exp(i\ell[\theta_p - \theta_k])}{i((m \pm Q)\omega) - i(\ell \pm \nu)\omega_1 + 0^+} \quad (3)$$

where  $0^+$  means take the limit as this term goes to zero though positive numbers. When pickup noise is included the equation for the transverse emittance of particles with revolution frequency  $\omega$  is

$$\frac{d\epsilon_\omega}{dt} = S_\omega \epsilon_\omega + \sum_{m, \pm Q} \frac{\pi N}{\omega_0} \left| \frac{G[(m \pm Q)\omega]}{\epsilon[(m \pm Q)\omega]} \right|^2 \epsilon_{\text{noise}, m \pm Q} \quad (4)$$

For low gain and resolution bandwidth larger than  $\omega_0$  the signal to noise ratio near frequency  $\omega_0(m \pm Q)$  is  $\epsilon_{rms}/\epsilon_{\text{noise}, m \pm Q}$  with  $\epsilon_{rms} = \int f(\omega) \epsilon_\omega d\omega$  the rms emittance. Changing the gain downstream of the noise source does not change  $\epsilon_{\text{noise}, m \pm Q}$ . As cooling proceeds  $\epsilon_{\text{noise}, m \pm Q}$  remains fixed and the signal to noise ratio drops.

**Longitudinal cooling** Taking  $x = E - E_0$  as energy variable with no measurement noise the equation of motion for particle  $k$  is [5]

$$\dot{x}_k = \sum_{j=1}^N \sum_{n=-\infty}^{\infty} G(n\omega(x_j), x_j) e^{in(\theta_k - \theta_j)} \quad (5)$$

where  $N$  is the number of particles in the ring,  $G(\Omega, x) = G_F(\Omega) + xG_P(\Omega)$  with  $G_F(\Omega)$  the filter cooling gain and  $G_P(\Omega)$  the electrical part of the Palmer cooling gain. A damped diffusion equation can be obtained,

$$\frac{\partial f(x, t)}{\partial t} = -\frac{\partial}{\partial x} \left[ F(x, t)f - D(x, t)\frac{\partial f}{\partial x} \right] \quad (6)$$

where

$$F(x, t) = \sum_m \frac{G(m\omega(x), x)}{\epsilon_L(m\omega(x))} e^{im(\theta_p - \theta_k)}$$

In the notation of [5], section 4,  $G_m(x) = G(m\omega(x), x)e^{im(\theta_p - \theta_k)}$ . When summed over all revolution lines the dielectric response is purely a function of frequency,

$$\begin{aligned} \epsilon_L(\Omega) &= 1 + \sum_m \int dx N \frac{\exp[i m(\theta_p - \theta_k)]}{i\Omega - im\omega(x) + 0^+} \\ &\times G(\Omega, x) \frac{\partial}{\partial x} (\omega(x)f(x, t)) \end{aligned} \quad (7)$$

The diffusion coefficient is

$$\begin{aligned} D(x, t) &= \sum_m \frac{N\pi}{|m|} \left| \frac{dx}{d\omega} \right| \left| \frac{G(m\omega(x), x)}{\epsilon_L(m\omega(x))} \right|^2 f(x, t) \\ &+ \sum_m N\pi \left| \frac{G(m\omega(x), x)}{\epsilon_L(m\omega(x))} \right|^2 U_m \end{aligned} \quad (8)$$

where  $U_m$  is the ratio of noise to signal power measured with a large resolution bandwidth.

**Longitudinal stacking** Solutions of Eq.(6) with  $\partial f/\partial t = 0$  approximate the central part of the solution during the accumulation of antiprotons. For no measurement noise and  $\epsilon_L = 1$  one has

$$\begin{aligned} -f(x) + \sum_m G_m(x) f \frac{\partial f}{\partial x} \sum_m \frac{N\pi}{|m|} \left| \frac{dx}{d\omega} \right| |G_m(x)|^2 \\ = \Psi_0 \end{aligned} \quad (9)$$

where  $\Psi_0$  is the flux of cooling particles. Setting  $G_m(x) = \alpha/f(x)$  for  $m_- < |m| \leq m_+$  and maximizing the cooling with respect to  $\alpha$  one finds

$$f(x) = f(x_1) \exp \left\{ \frac{(m_+ - m_-)^2 d\omega/dx}{2\pi N \ln \left( \frac{m_+}{m_-} \right) \Psi_0} (x - x_1) \right\} \quad (10)$$

where  $x_1$  is an energy in the constant flux range.

**Bunched beam cooling** For rough estimates one can use Eq.(1) with  $N$  corresponding to the number of particles in the ring that would create a current equal to the peak beam current. The major unknown is the mixing factor,  $M$ . With large synchrotron sideband overlap, as occurs in

full rf buckets,  $M$  is the ratio of peak Schottky power to average Schottky power, as in coasting beams [11]. Mixing from intra-beam scattering (IBS, Sec.2.4.12) is important in the FNAL recycler [12]. For higher accuracy the damped diffusion equations for bunched beams are much more complicated than those for coasting beams [13] though calculations without synchrotron sideband overlap [13], or neglecting signal shielding [11], have been done. For application to RHIC it was found that multiparticle simulations proved both fast and reliable [9]. To simulate  $N_r$  real particles using  $N_s$  simulation particles one simply multiplies the real gain by  $N_r/N_s$  and tracks for  $N_s/N_r$  fewer turns [14]. The relevant algorithms are identical to those used for the simulation of coherent instabilities. As an example consider longitudinal cooling. The first update will take place at the rf cavity and, since synchrotron tunes are small compared to one, we may place the stochastic cooling kicker at the same spot,

$$\begin{aligned} \bar{x}_n &= x_n + qV(\tau_n) \\ \bar{\tau}_n &= \tau_n + \kappa_1 \bar{x}_n \end{aligned}$$

where  $x_n = E - E_0$  for particle  $n$ ,  $\tau_n$  is the arrival time of particle  $n$  with respect to the synchronous particle, and the bars denote updated variables. The parameter  $\kappa_1 = f\eta T_{\text{rev}}/\beta^2 E_0$  accounts for particle slip when traversing a fraction of the ring  $f$  between kicker and pickup. At the pickup one accumulates the line density array for  $k = 1, 2, \dots, N_{\text{grid}}$ ,

$$\lambda(t_k) = \sum_{n=1}^{N_s} \hat{\delta}(t_k - \bar{\tau}_n)$$

where  $\hat{\delta}(t)$  is the triangle function for linear interpolation on the grid  $t_k = k\Delta t$ . Next one transports from the pickup to the kicker via  $\bar{\tau}_n = \bar{\tau}_n + (1-f)\kappa_1 \bar{x}_n$ . The kick for simple filter cooling starts with  $\Delta\lambda(t_k) = \lambda(t_k) - \lambda_{\text{old}}(t_k)$  where the array  $\lambda_{\text{old}}$  was accumulated on the previous turn. Then  $\Delta\lambda(t)$  is Fourier transformed, multiplied by the frequency dependent gain, and inverse transformed. The voltage obtained is added to the rf voltage and the update sequence is repeated.

Additional effects such as betatron coupling and IBS are easily included, though one must be careful not to make  $N_s$  too small and introduce unphysical mixing due to the increased IBS rates. The results with several competing factors operating can be quite different from the estimates of individual effects. For instance, longitudinal action

diffusion from IBS turns out to be very important for transverse cooling in RHIC. This is because particles with small synchrotron amplitude have poor cooling [13, 14] but longitudinal diffusion causes all particles to have significant synchrotron amplitudes at least some of the time, leading to even cooling.

### 2.7.1.2 Hardware

The cooling system is a wide-band feedback loop with bandwidths of a few hundred MHz to a few GHz. Pickups are generally wide-band devices and employ multiple slots to extract sufficient beam power. The individual slot signals can be combined using a meander line [7]. Alternately, the coupling slots can modify the phase velocity of the waveguide mode leading to resonant buildup over a wide frequency range [15]. Both of these technologies require careful design but yield wide band high signal to noise devices. The very strong Schottky signals of the gold ions in the RHIC collider allowed the longitudinal pickup to be very simple. A pair of commercial waveguide launchers on either side of the beam pipe are combined in sum mode. A ceramic window keeps all the electronic components outside the vacuum, resulting in an inexpensive, robust design.

During design both the Panofsky-Wenzel theorem and the potential theorem [16] significantly augment the usefulness of electromagnetic simulation codes. Let  $s = z + vt$  denote the longitudinal position of the kicked particle as a function of time. For an isolated kicker the Panofsky-Wenzel theorem (Sec.3.2.1) states there exists a function  $\Phi(x, y, z)$  such that the momentum kick is  $\Delta(p_x, p_y, p_z) = \partial\Phi/\partial\mathbf{r}$  where  $\mathbf{r} = (x, y, z)$ . The potential theorem states that  $\Phi$  obeys

$$\frac{\partial^2\Phi}{\partial x^2} + \frac{\partial^2\Phi}{\partial y^2} + \frac{1}{\gamma^2}\frac{\partial^2\Phi}{\partial z^2} = 0 \quad (11)$$

where  $\gamma$  is the Lorentz factor. For coasting beam cooling the front end noise in the pickup is a serious concern. Another difficulty is maintaining the appropriate gain and phase functions over very wide bandwidth. Various sorts of equalizers are often needed and feedback on delays is commonplace. For bunched beams there is the additional complication of beam driven coherent lines [17]. In RHIC, the signal power from coherent lines is often 100 times larger than the Schottky power. A large dynamic range is needed to keep intermodulation products under control.

Longitudinal filter cooling usually employs one turn delays between pickup and kicker. The simplest and most common type is of the form  $S_1(t) = S_0(t) - S_0(t - T_{rev})$  where  $S_0(t)$  is the pickup signal,  $T_{rev}$  is a highly accurate one turn delay and  $S_1(t)$  is the output signal. Single mode optical fibers allow for multi-GHz bandwidths and are inexpensive. Careful temperature control is required to keep length variations well under 1/4 wavelength and various sorts of feedback on the delay are standard.

Kickers for coasting beams look much like pickups, as expected from the Lorentz reciprocity theorem. One simply reverses the beam direction and uses the same solution to Eq.(11) [16]. For bunched gold beams in RHIC, longitudinal cooling requires an rms voltage of 3 kV. Fortunately, the bunch spacing of 107 ns is much larger than the 5 ns bunch length. This allowed for a Fourier series based approach using resonant cavities to generate the kick [9, 14].

### 2.7.1.3 New techniques

Microwave stochastic cooling systems are limited to bandwidths of a few GHz. While this is adequate for condensing antiprotons and other rare particles or cooling the gold beams in RHIC it is not useful for cooling high density proton beams in colliders. For such beams bandwidths of several hundred GHz are required. Two technologies have been suggested for cooling such beams.

Optical stochastic cooling employs a wiggler pickup, optical amplifier, and wiggler kicker to close the cooling feedback loop [18, 19]. Operations at micron wavelengths with 10% bandwidth are envisioned. The bandwidth is of order  $10^{13}$  Hz but there are significant challenges. The short wavelengths require linear but nearly achromatic optics and the laser amplifiers push the state of the art. For a laser power limited system the energy cooling rate is [19]

$$\frac{d\sigma_E^2}{dt} = -\frac{\sigma_E^2}{n_E T_{rev}} \text{ with } \frac{1}{n_E^2} \approx \frac{\bar{P}cZ_0K^2}{\bar{I}/q\lambda\sigma_E^2/q^2} \quad (12)$$

where  $\sigma_E$  is the rms energy spread,  $q$  is the charge of the ion,  $\bar{P}$  is the average laser power,  $\bar{I}$  is the average beam current,  $K = qcB\lambda_u/(2\pi Mc^2)$  is the undulator parameter,  $Z_0 = 377\Omega$ , and  $\lambda$  is the central wavelength for the laser. The maximum laser power is  $P_{max} = \bar{P}I_{max}/\bar{I}$  where  $I_{max}$  is the peak beam current.

## Sec.2.7: BEAM COOLING

Coherent electron cooling [20] involves an electron bunch comoving with the hadron bunch. In the modulator the hadrons induce a density modulation on the electrons. The modulation is then amplified in a high gain free electron laser. In the kicker the electron bunch is again merged with the hadron bunch and the electrons kick the hadrons. Appropriate optics and timing result in cooling the hadrons. Central wavelengths of order 10 microns are envisioned and appropriate optics are being designed. High energy, high current energy recovery linacs appear capable of supplying the electron beams to cool proton colliders. The energy cooling rate is [20]

$$\frac{1}{n_E} \approx \frac{4\pi G_L r_c L_k D}{\lambda \gamma \sigma_{x,e}^2} (1 - \cos \psi_p) \frac{\sin \psi_k}{\psi_k} \quad (13)$$

where  $G_L$  is the gain of the free electron laser,  $r_c = q^2/4\pi\epsilon_0 mc^2$  is the classical particle radius,  $L_k$  is the length of the kicker,  $D = d\ell/d(\Delta p/p)$  is the path length variation with momentum for the ions,  $\gamma$  is the Lorentz factor,  $\sigma_{x,e}$  is the rms radius of the electron beam,  $\psi_p \approx \pi/2$  is the plasma phase advance in the pickup and  $\psi_k \approx \pi/2$  is the plasma phase advance in the kicker. It has been assumed that the light wavelength,  $\lambda$  is long enough,  $\lambda/2\pi \geq D\Delta p/p$ , so that all ions are cooled.

## References

- [1] S. van der Meer, CERN/ISR-PO/72-31 (1972); S. van der Meer, RMP 57 (1985) 689
- [2] G. Caron et al, PAC 79, p.3456; G. Caron et al, PL B77 (1978) 353
- [3] G. Lambertson et al, Proc. Int. Conf. on High Energy Acc. (1980) p.794; R.L. Hogrefe et al, PAC 81, p.2455
- [4] D. Mohl et al, Phys. Rep. 58 (1980) 73
- [5] J. Bisognano, C. Leemann, AIP Proc. 87, p.583 (1981)
- [6] John Marriner, David McGinnis, AIP conf. 249 p.693 (1992)
- [7] See e.g. S. Nagaitsev, R.J. Pasquinelli eds., AIP conf. 821 (2005); COOL07
- [8] R. Pasquinelli, PAC95, p.2379
- [9] M. Blaskiewicz, J.M. Brennan, F. Severino, PRL 100 174802 (2008)
- [10] M. Blaskiewicz et al, IPAC11, p.913 (2011)
- [11] J. Wei, CERN 94-03 (1994)
- [12] D. Broemelsiek et al, PRST-AB **8**, 124404 (2005)
- [13] S. Chattopadhyay, LBL-14826 (PhD thesis) (1982); also several papers in PAC83
- [14] M. Blaskiewicz, J.M. Brennan, COOL07
- [15] D. McGinnis, PAC99, p.1713
- [16] G. Lambertson, AIP Proc. 153 (1987) p.1414
- [17] D. Boussard, Lecture Notes on Phys. 296, p.289
- [18] A.A. Mikhailichenko, M.S. Zolotorev, PRL 71, p.4146 (1993).
- [19] M.S. Zolotorev, A.A. Zholtens, PR E50, p.3087 (1994)
- [20] V.N. Litvinenko, Y.S. Derbenev PRL 102, 114801 (2009)

### 2.7.2 Electron Cooling

*S. Nagaitsev, FNAL*

Electron cooling, which has become a conventional tool in accelerator technology, makes it possible to increase or to control the phase space density of heavy charged particle (protons, ions, antiprotons) beams in a storage ring. About ten such storage rings, referred to as coolers, are presently operating in Europe, US, China and Japan. These have greatly expanded the experimental capabilities in the study of elementary particles, nuclear and atomic physics.

The initial idea belongs to Budker [1] and is based on the model of two-component electron-ion plasma. A monochromatic electron beam is merged with a straight section of the storage ring. The average longitudinal velocities of the ion and electron beams are made equal and the two beams are made to overlap transversely, thus creating electron-ion plasma co-propagating for a period of time until the electron beam is separated from the ion beam and discarded. The ions, while traveling with electrons, undergo Coulomb scattering and, under certain conditions, transfer some of their thermal energy to electrons. Ultimately, in the absence of other heating mechanisms, the ion temperature in the beam frame (BF) of reference becomes equal to the effective electron temperature. As a result, the ion beam angular spread,  $\theta_i$ , becomes substantially smaller than the electron beam angular spread,  $\theta_e$ ,

$$\theta_i = \frac{p_\perp}{p_0} \approx \sqrt{\frac{m}{M}} \theta_e \quad (1)$$

where  $m$  and  $M$  are the electron and ion masses respectively. Early theoretical investigations of the electron cooling process [2] demonstrated the importance of achieving low electron beam temperature (angular spread).

**Friction force** In a simplified electron-ion plasma model, an ion traveling through the electron “gas” experiences a friction force

$$\mathbf{F}(\mathbf{v}) = 4\pi n_e m (Zr_e c^2)^2 \Lambda \int_{-\infty}^{+\infty} f(v_e) \frac{\mathbf{v}_e - \mathbf{v}}{|\mathbf{v}_e - \mathbf{v}|^3} d^3 v_e \quad (2)$$

with  $n_e$  the BF electron density,  $\Lambda$  is the Coulomb log ( $\approx 5-10$ ) and  $f(v_e)$  is the electron velocity distribution function in the BF. For example, an anisotropic Gaussian distribution function has

$$f(\mathbf{v}_e) = \frac{1}{(2\pi)^{\frac{3}{2}} \sigma_{\perp}^2 \sigma_{\parallel}^2} \exp \left[ -\frac{v_{e\perp}^2}{2\sigma_{\perp}^2} - \frac{v_{e\parallel}^2}{2\sigma_{\parallel}^2} \right] \quad (3)$$

with  $\sigma_{\perp}$  and  $\sigma_{\parallel}$  the BF transverse and longitudinal rms velocity spreads, which can be expressed through laboratory-frame (LF) quantities as

$$\sigma_{\perp} = \beta\gamma\theta_e c, \quad \sigma_{\parallel} \approx \frac{\delta E}{\beta\gamma mc} \quad (4)$$

where  $\theta_e$  is the LF rms electron beam angular spread and  $\delta E$  is the LF rms energy spread of the electron beam. Typically  $\sigma_{\perp} \gg \sigma_{\parallel}$  leading to what is called a “flattened” electron distribution [3]. The electron beam angular spread is mainly determined by (a) the cathode temperature, (b) the electron beam space charge and (c) the transport line imperfections. The electron energy spread has three major contributions: (a) the energy ripple, (b) the multiple-Coulomb scattering, and (c) the electron beam density fluctuations (aka the longitudinal-longitudinal relaxation).

Approximate expressions for the cooling force (2) can be found in [3, 4]. Consider two simplified cases when an ion has only one component of its velocity vector either longitudinal or transverse:  $v_{\parallel} \ll \sigma_{\parallel}$  and  $v_{\perp} \ll \sigma_{\perp}$ . We have ( $\sigma_{\perp} \gg \sigma_{\parallel}$ )

$$F_{\parallel} \approx -4\sqrt{2\pi} n_e m (Zr_e c)^2 \Lambda \frac{v_{\parallel} c^2}{\sigma_{\perp}^2 \sigma_{\parallel}} \quad (5)$$

$$F_{\perp} \approx -\pi\sqrt{2\pi} n_e m (Zr_e c)^2 \Lambda \frac{v_{\perp} c^2}{\sigma_{\perp}^3} \quad (6)$$

**Cooling rates** In these cases the friction forces (5) and (6) are proportional to the ion velocity and thus the particle loses its thermal energy in the BF exponentially with a time constant

$$\tau_{\parallel}^{-1} = 4\sqrt{2\pi} n_e (Zr_e)^2 c \Lambda \frac{m}{M} \frac{c^3}{\sigma_{\perp}^2 \sigma_{\parallel}} \quad (7)$$

$$\tau_{\perp}^{-1} = \pi\sqrt{2\pi} n_e (Zr_e)^2 c \Lambda \frac{m}{M} \frac{c^3}{\sigma_{\perp}^3} \quad (8)$$

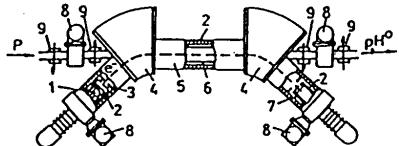


Figure 1: Schematic diagram of the NAP-M electron cooling system: 1 – electron gun, 2 – solenoids, 3 – anodes of the electron gun, 4 – toroids, 5 – cooling section, 6 – vacuum chamber, 7 – electron beam collector, 8 – vacuum pumps, and 9 – dipole correctors.

To express these values through LF quantities one has to use proper Lorentz transformations and to recall that the cooling section occupies only a fraction,  $\eta$ , of the ring circumference. Thus we obtain the following cooling rates:

$$\lambda_{\parallel} = 4\sqrt{\frac{2}{\pi}} \eta \Lambda \frac{I_e}{e} \left( \frac{Zr_e}{a} \right)^2 \frac{m}{M} \frac{mc^2}{\beta^2 \gamma^3 \theta_e^2 \delta E} \quad (9)$$

$$\lambda_{\perp} = \sqrt{2\pi} \eta \Lambda \frac{I_e}{e} \left( \frac{Zr_e}{a} \right)^2 \frac{m}{M} \frac{1}{\beta^4 \gamma^5 \theta_e^3} \quad (10)$$

where  $I_e$  and  $a$  are the electron beam current and radius. A uniform electron density distribution is assumed. A numerical example: consider 100 MeV protons being cooled by a 54.5-keV, 1-A electron beam with a 1 cm radius and assume that the electron beam rms energy spread is 10 eV and its rms angular spread is 2 mrad. Suppose that  $\eta = 0.02$  and  $\Lambda = 10$ . The cooling times are 0.13 and 4.5 s correspondingly. For large ion beam angular spreads,  $\lambda \propto (\theta_i)^{-3}$ .

To obtain a more complete picture of the cooling process one needs to average Eq.(2) over the ion beam distribution, taking into account both the betatron and synchrotron motion.

**Magnetic field** Electron cooling was first realized in 1974 in experiments on cooling 65-MeV protons at the NAP-M storage ring at BINP [5]. Fig.1 shows the schematic of the system.

An electron gun was placed into a solenoid producing the longitudinal guiding magnetic field, which accompanies the beam until it reaches the collector. The longitudinal magnetic field enabled passing an electron beam over the required, rather long, (a few meters) distance without dilution. Moreover, it allowed to form an intense electron beam with a small angular spread. The longitudinal magnetic field transforms the electron motion, perpendicular to it, into gyrations about the magnetic field lines. Cooling by “magnetized” electrons differs substantially from the simplified

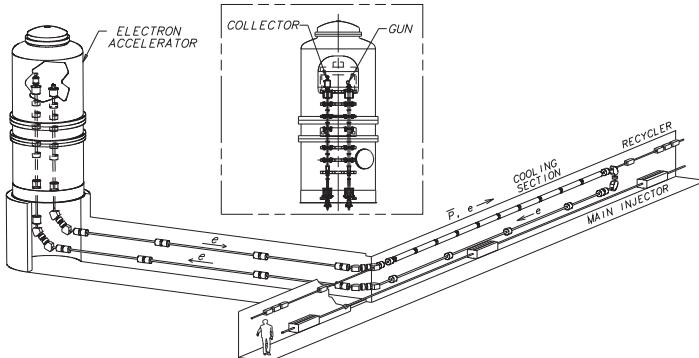


Figure 2: Schematic layout of the Recycler electron cooling system and accelerator cross-section (insert).

description of Eq.(2): it makes electron cooling faster (at small relative velocities) and the equilibrium beam temperatures lower. The reason is as follows. The Larmor radius of the electron gyration can be smaller than the impact parameter of the electron – ion collision. This makes the electron interact with an ion longer and thus increases the friction force substantially. The effect was first discovered experimentally at NAP-M [6] and then studied theoretically [7, 8, 9]. A dedicated experimental study of this effect was performed at BINP with a single-pass setup MOSOL [10]. A factor of  $> 100$  enhancement in the cooling force at small relative velocities was observed. These studies lead to the realization that quality (field line straightness) of the solenoidal field is critical to fast magnetized cooling. A typical field uniformity achieved in state-of-art electron coolers is  $\Delta B_{\perp}/B < 10^{-4}$  at a field level of  $B \sim 1$  kG.

**Practical realizations** Following the NAP-M success, a number of coolers were built in 1970-90 [3, 11, 12] with key features similar to that of the first cooler Fig.1. The required electron beam currents of  $\sim 1$  A and beam energies of up to 300 keV demanded a beam recirculation scheme in which the electron beam is decelerated in the collector, giving most ( $> 99\%$ ) of its energy back to the high-voltage power supply. In practice, this allowed achieving very low energy ripple ( $< 100$  ppm) and acceptable vacuum gas loads in the cooling section. A description of the recently built LEIR electron cooler can be found in [13].

**Relativistic electron cooling** As of now, Fermilab's Electron Cooler [14] remains unique in its ability to cool relativistic (8 GeV) antiprotons,

exceeding in energy of “low-energy” coolers by more than an order of magnitude. Fig.2 shows the schematic layout of the Fermilab system. The dc electron beam is generated by a thermionic gun, located in the high-voltage terminal of the electrostatic (Van-de-Graaff type) accelerator. This accelerator is incapable of sustaining dc beam currents to ground in excess of about  $100 \mu\text{A}$ . To attain the electron dc current of 500 mA, a recirculation scheme is employed. A typical inefficiency of such a process is 20 ppm.

The Fermilab system employs a unique beam transport scheme [15]. The electron gun is immersed in a solenoidal magnetic field, which creates a beam with large angular momentum. After the beam is extracted from the magnetic field and accelerated to 4.3 MeV, it is transported to the 20-m long cooling section solenoid using conventional focusing elements. The cooling section solenoid removes this angular momentum and the beam is made round and parallel such that the beam radius,  $a$ , produces the same magnetic flux,  $Ba^2$ , as at the cathode. The magnetic field in the cooling section is quite weak (100 G) and the kinetics of the electron-antiproton scattering is weakly affected by the magnetic field.

**Applications** In elementary particle physics, the most important application of electron (as well as stochastic) cooling is in accumulation and storage of antiprotons destined for the colliding beam experiments. In nuclear physics the appearance of coolers has revealed qualitatively new possibilities: precision experiments with extremely small (down to  $10^{-6}$ ) relative energy spreads, thin internal targets, mass spectrometry of super-high resolution and the separation and storage of long lived

isotopes and isomers [16]. Highlights include using electron cooling to prepare dense bunches of lead ions for ion-ion collisions at the LHC [13] and very demanding cooling schemes for the FAIR project at GSI [17]. In atomic physics the electron beam can be used as the “target”, allowing for a precision study of ion recombination. Finally, electron cooling has led to the development of ordered (crystalline) ion beams [18, 19].

## References

- [1] G.I. Budker, At. Energy 22, 346 (1967) [Sov. Atomic Energy 22, 438 (1967)]
- [2] Ya.S. Derbenev, A.N. Skrinsky, PA 8, 1 (1977)
- [3] I.N. Meshkov, Phys. Part. Nucl. 25(6), 631 (1994)
- [4] H. Poth, Phys. Reports 196 (1990) 135
- [5] G.I. Budker et al, PA 7, 197 (1976)
- [6] G.I. Budker et al, USSR High Energy Accel. Conf., 1976; translation CERN/PS/DL/Note 76-25
- [7] Ya.S. Derbenev, A.N. Skrinsky, PA 8, 235 (1977)
- [8] M. Bell, PA 10, 101 (1980)
- [9] V.V. Parkhomchuk, NIM A 441, 9 (2000)
- [10] N.S. Dikansky et al, Preprint INP 88-61, 1988
- [11] M. Bell et al, NIM 190, 237 (1981)
- [12] T. Ellison et al, IEEE Trans. Nucl. Sci. NS-30, 2370 (1983)
- [13] G. Tranquille et al, COOL07
- [14] S. Nagaitsev et al, PRL 96, 044801 (2006)
- [15] A. Burov et al, PRST-AB, 094002 (2000)
- [16] H.-J. Kluge et al, NIM A 532, 48 (2004)
- [17] <http://www.gsi.de/fair/>
- [18] M. Steck et al, PRL 77, 3803 (1996)
- [19] H. Danared et al, PRL 88, 174801 (2002)

### 2.7.3 Laser Cooling

*J. Hangst, CERN  
A. Noda, Kyoto U.*

Laser cooling [1], where an ion is excited by absorption of a photon and returns back to the ground state by spontaneous emission of a photon, has been successfully applied for trapping ions and cooling them to a temperature below 1 mK [2]. A moving beam in a storage ring is typically hot and its constituent particles have a large fluctuation in kinetic energies in the CM system. This fluctuation can be reduced by absorption and spontaneous emission of the photons. As shown in Fig.1, an ion in a lower energy state, A, moving with a velocity  $v_0$  in a certain direction, will get excited to an upper energy state, B,

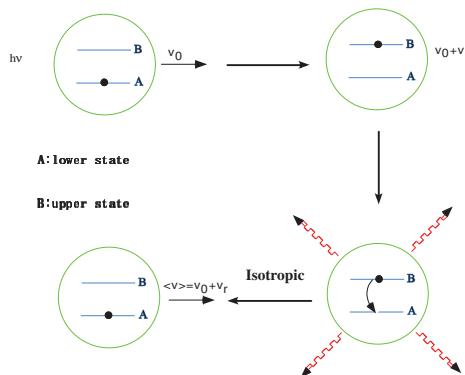


Figure 1: Mechanism of absorption and spontaneous emission of photons in a laser cooling process.

by irradiation of the co-propagating or counter-propagating laser with the Doppler shifted frequency,  $\nu$ , satisfying such a condition as  $h\nu$  coincides with the band gap between upper state, B and lower state, A. The ion obtains the momentum of  $h\nu/c$  in the direction of laser irradiation by absorption of the photon. Thus the ion is slightly accelerated (co-propagating case) or decelerated (counter-propagating case). If the lifetime of the excited state is short enough, the ion immediately returns back to the ground state by emitting a photon. Such a spontaneous emission process occurs isotropically and recoil momentum given to the ion is statistically zero. So after the excitation and de-excitation processes Fig.1, the ion on the average only receives a small velocity gain in the laser direction keeping other conditions unchanged. In such a case as a co-propagating laser is irradiated and the laser frequency is at first detuned to red (longer wavelength) side and then swept toward the resonant frequency, the velocity distribution of the ion can be modified as shown in Fig.2 creating a sharp peak in the velocity distribution.

The laser cooling force of a single laser in its propagation direction is [3]

$$\vec{F}_1(\vec{v}, \vec{k}) = \frac{\hbar \vec{k}}{2} ST \frac{(\frac{\Gamma}{2})^2}{(\Delta - \vec{v} \cdot \vec{k})^2 + (\frac{\Gamma}{2})^2(1+S)} \quad (1)$$

where  $\hbar$ ,  $\Gamma = 1/\tau$ ,  $\Delta = \omega - \omega_0$ ,  $\vec{v}$ ,  $\vec{k}$  and  $S$  are reduced Planck constant, the natural linewidth of the cooling transition, the angular frequency detuning of the laser relative to the resonance, ion velocity, the wave vector of the laser and the saturation parameter,  $S = 2(\frac{\mu E}{\hbar \Gamma})^2$  ( $E$ : laser electric field,  $\mu$ : dipole moment for the transition).

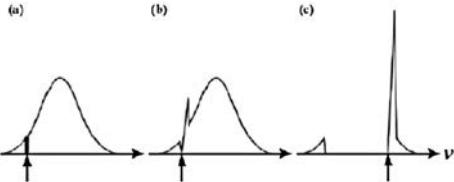


Figure 2: (a) Irradiation of co-propagating laser, results in acceleration of the ions in the velocity range. (b) The ion is given the energy as large as  $h\nu$  from the laser and corresponding velocity change of the ion occurs. (c) After sweeping the laser frequency through the whole velocity range of the ion beam, ions are collected in a sharp velocity range.

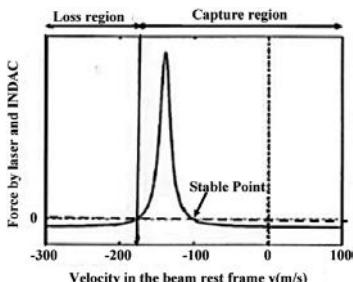


Figure 3: Total cooling force superposing the laser cooling force and the induction voltage by INDAC. Stable point and capture region are indicated.

For laser cooling with a single laser, a counteracting force to the above laser cooling force is needed in order to keep the ion beam at a certain stable orbit. Such a force is provided with an induction accelerator or rf acceleration voltage for the cases of coasting or bunched beams, respectively. In Fig.3, total cooling force created by combination of a single laser and a counteracting deceleration voltage applied through an Induction Accelerator (INDAC) is given together with an indication of stable point.

For two counter propagating lasers with the same intensity and detuning, the total cooling force is given by superposition (Fig.4),

$$\vec{F}_2(\vec{v}, \vec{k}) = \vec{F}_1(\vec{v}, \vec{k}) + \vec{F}_1(\vec{v}, -\vec{k}) \quad (2)$$

In the small velocity region, the force can be approximated by a linear function of the relative velocity,

$$F = m \frac{dv}{dt} \approx -\beta v \quad (\beta = \frac{dF}{dv}|_{v=0}) \quad (3)$$

From Eq.(3), the velocity  $v$  is known to be represented as an exponential function of  $-(\beta/m)t$

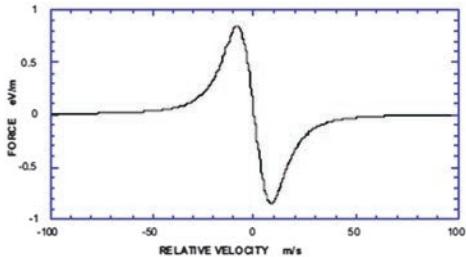


Figure 4: Total cooling force given by a combined system of two lasers co- and counter-propagating with the ion beam.

Table 1: Main parameters of laser cooling of moving ions.

| Ion Species                      | ${}^7\text{Li}^+$        | ${}^9\text{Be}^+$ | ${}^{24}\text{Mg}^+$ |
|----------------------------------|--------------------------|-------------------|----------------------|
| Lower level                      | $2s^3S_1(F=\frac{5}{2})$ | $2s^2S_{1/2}$     | $3s^2S_{1/2}$        |
| Upper level                      | $2p^3P_2(F=\frac{7}{2})$ | $2p^2P_{3/2}$     | $3p^2P_{3/2}$        |
| Life time (L)                    | 50 s                     | stable            | stable               |
| Life time (U)                    | 43 ns                    | 8.2 ns            | 3.7 ns               |
| Line width $\frac{\Gamma}{2\pi}$ | 3.7 MHz                  | 19.3 MHz          | 42.6 MHz             |
| Wave length                      | 548 nm                   | 313 nm            | 280 nm               |
| Doppler limit                    | 89 $\mu\text{K}$         | 440 $\mu\text{K}$ | 1 mK                 |

and the corresponding damping time is  $\sim 10^{-6}$  s for such ion beams as  ${}^7\text{Li}^+$ ,  ${}^9\text{Be}^+$  and  ${}^{24}\text{Mg}^+$ , ever laser cooled. The equilibrium temperature of laser cooling is determined by the balance of the cooling force and other heating mechanisms. Without external heating, equilibrium is established between the cooling force and the spontaneous emission, and is known as the Doppler limit,

$$T_D = \frac{\hbar\Gamma}{2k_B} \quad (4)$$

where  $k_B$  is the Boltzmann constant ( $8.62 \times 10^{-5}$  eV/K).

The Doppler limit for laser cooled ions in a storage ring is listed in Tab.1 together with other main parameters. They are in the range of  $10^{-4} \sim 10^{-3}$  K.

**Experiments** Laser cooling has been first applied to a moving beam at the TSR, MPI-K, Heidelberg for coasting  ${}^7\text{Li}^+$  ion beam [4]. Its metastable  $2s^3S_1(F=5/2)$  state is excited to the upper level  $2p^3P_2(F=7/2)$  with a mean radiative lifetime of 43 ns, which has a branching ratio less than  $10^{-5}$  to the ground state  ${}^1S_0$ . The lower metastable  ${}^3S_1$  state has a radiative lifetime of 50 s, long enough compared to ion storage life time  $\sim 2$ s. Thus, the  ${}^3S_1(F=5/2) \rightarrow {}^3P_2$

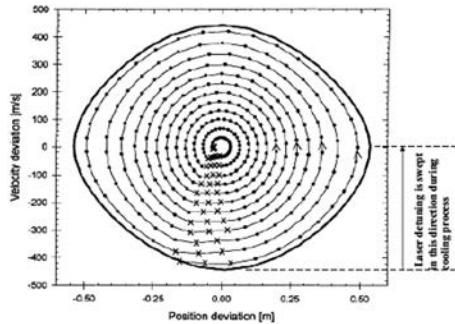


Figure 5: Scheme of bunched beam laser cooling. Laser frequency detuning is swept from the bottom of the separatrix to the center of resonance condition.

( $F=7/2$ ) transition approximately plays the role of two level system. Ring dye lasers co- and counter-propagating with the ion beam with frequencies which coincide with the level difference of 548.5 nm after Doppler shift, are utilized for excitation of the ion, resulting in the longitudinal temperature of 3 K. Laser cooling is applied also to  ${}^9\text{Be}^+$  and  ${}^{24}\text{Mg}^+$  ions at TSR and ASTRID, IPA in Aarhus for coasting and bunched beams, respectively [5, 6]. Fig.5 illustrates the scheme of bunched beam laser cooling [6]. The laser frequency is red shifted to the lowest energy level covered by the rf separatrix and then the laser frequency is shifted to the central value of the separatrix according to the cooling requirement. The equilibrium temperature of the cooled beam is determined by the balance between the laser cooling force and the heating process due to intra-beam scattering, which limits the attained beam temperature around 1K for intense beam.

**3-D cooling** Concerning longitudinal cooling, a beam temperature less than 1 mK can be attained for a small fraction ( $\sim 10^{-4}$ ) of  ${}^7\text{Li}^+$  beam, which is in a metastable  ${}^3\text{S}_1$  state [7] and beam temperature lower than 30 mK is attained for  ${}^9\text{Be}^+$  beam which is pre-cooled by electron cooling [5]. Attainable equilibrium temperature depends on the ion beam density.

In order to extend the laser cooling force to 3-D, various approaches have been applied. At TSR, transverse laser cooling with the use of intrabeam scattering has been demonstrated for  ${}^9\text{Be}^+$  ion beam [8]. The transverse cooling force, however, depends on the passive energy transfer between transverse and longitudinal directions by

IBS and cooling efficiency is not so good to approach a cold beam. Another scheme to utilize longitudinal-horizontal coupling through ring dispersion (transverse laser cooling through dispersive coupling) has also been demonstrated at TSR. This acts as a single-particle effect independent of the ion density in contrast to IBS cooling. The transverse cooling rate realized, however, is  $\sim 1 \text{ s}^{-1}$ , still far below typical longitudinal laser cooling rates [9].

As an efficient way of horizontal and longitudinal coupling, a scheme to utilize an rf cavity located at a dispersive position in the ring has been proposed. The synchrotron motion and horizontal betatron oscillation are coupled by a Synchro-betatron resonance [10]. In order to couple the horizontal and vertical directions, the usual horizontal-vertical coupling near a difference resonance is applied using a solenoid or a skew quadrupole magnet. The synchro-betatron coupling scheme is now under experimental research at S-LSR, ICR, Kyoto Univ. and experimental indication of such a resonant coupling between two degrees of freedom, longitudinal an horizontal directions, has been obtained [11]. In this experiment, the optical measurement system of the transverse beam profile by detection of the laser-induced fluorescence (LIF) light with a CCD camera basically developed at ASTRID [12] played an essential role. As shown in Fig.6, S-LSR is designed to have a higher symmetry of 6 in order to satisfy the *maintenance condition* to keep beam crystal structure stable:

$$\nu_{h,v} \leq \frac{N}{2\sqrt{2}}$$

where  $\nu_{h,v}$  and  $N$  are betatron tunes in horizontal and vertical directions and number of superperiodicity of the ring [13]. As illustrated in Fig.7, S-LSR also provides a capability to keep the crystalline beam structure to be stable suppressing shear heating by superposing an electrostatic field,  $E$ , with the bending magnetic field,  $B$ , satisfying the relation

$$(1 + \frac{1}{\gamma^2})\vec{E} = -\vec{v}_0 \times \vec{B}$$

where  $\gamma$  and  $v_0$  are Lorentz factor and velocity of the beam, respectively [14, 15].

Circulating crystalline ion beam (Sec.2.7.5) has been realized with a circular RFQ storage ring, PALLAS at LMU, Munich, with application of laser cooling to  ${}^{24}\text{Mg}^+$  ion with the energy of

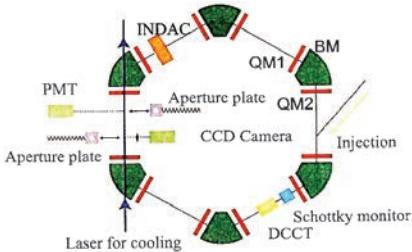


Figure 6: Layout of S-LSR. Six-fold symmetry is adopted for its lattice to satisfy the so-called maintenance condition.

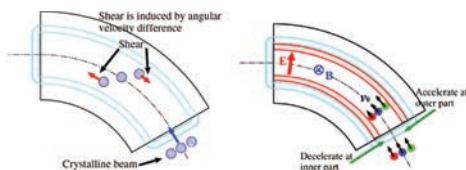


Figure 7: Illustration of shear heating and the deflection element combining the electrostatic and magnetic fields to suppress the shear heating adopted at S-LSR.

1 eV [16]. The crystal structure is found to be destroyed in case the ion energy is increased by acceleration to higher than 1 eV with a drift tube set inside the PALLAS due to shear heating.

Laser cooling of relativistic ions has also been performed at ESR in GSI, Darmstadt applying the Ar<sup>+</sup> ion laser with the wave length of 257.34 nm to the highly Doppler shifted C<sup>3+</sup> ions with the energy of 1.47 GeV (122 MeV/u) to excite the transition from 2S<sub>1/2</sub> to 2P<sub>1/2</sub> (transition wave length: 155.07 nm) and from 2S<sub>1/2</sub> to 2P<sub>3/2</sub> (transition wave length: 155.81 nm). Such a result of laser cooling is to be applied at the FAIR project [17].

**Perspective** Although the possibility to couple the transverse and longitudinal motions dynamically in order to “share” the laser cooling force among three dimensions [10] has been shown experimentally, its efficiency is not so high as be able to realize crystalline beam for the rather intense beam ( $\sim 10^7$  per ring) [11]. Experimental approach to increase the transverse cooling efficiency by reducing the ion beam intensity to the level as MD simulation predict realization of string beam ( $\sim 10^4$  per ring) [18] is under way which requires a more sensitive transverse profile detection system.

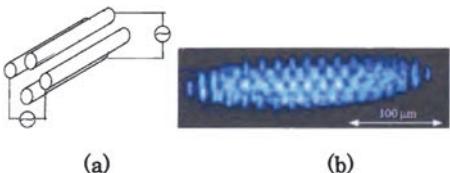


Figure 8: (a) Illustration of ion trapping with the rf electric fields at a linear ion trap, and (b) realized multi-layer ion crystal in a linear ion trap [21].

**Particle Trapping** Laser cooling has already been successfully applied to neutral atoms and charged ions trapped inside a trap, which has been utilized for spectroscopy, fundamental quantum physics, quantum computation etc. Neutral atoms are captured and trapped after laser cooling, in many cases, with the aid of magnetic field confining them by magnetic moment [19]. An ion trap is an apparatus for confining charged particles utilizing electromagnetic forces, e.g. Penning trap, utilizing static electric and magnetic fields, and rf trap, utilizing rf electric fields. Rf ion traps have configuration of 3-D quadrupoles (Paul Trap [20]) and a linear trap utilizing four rods with application of rf electric fields to confine charged ions in radial direction as shown in Fig.8(a), while confinement in the axial direction is attained by static potentials applied to end plates. With a linear ion trap, a crystalline beam of <sup>40</sup>Ca<sup>+</sup> has already been realized as shown in Fig.8(b) [21].

## References

- [1] T. Hänisch, A. Schawlow, Opt. Commun. 13, 68 (1975)
- [2] S. Chu et al, PRL 55, 48 (1985)
- [3] J. Gordon, A. Ashkin, PRA A21, 1606 (1980)
- [4] S. Schröder et al, PRL 64, 2901 (1990)
- [5] W. Petrich et al, PRA 48, 2127 (1993)
- [6] J.S. Hangst et al, PRL 74, 4432 (1995)
- [7] J.S. Hangst et al, PRL 67, 1238 (1991)
- [8] H.J. Meissner et al, PRL 77, 623 (1996)
- [9] I. Lauer et al., PRL 81, 2052 (1998)
- [10] H. Okamoto, A. Sessler, D. Möhl, PRL 72, 3977 (1994)
- [11] M. Nakao et al, to be published
- [12] N. Madsen et al, PRL 83 (1999) 4301
- [13] J. Wei, X.-P. Li, A. M. Sessler, PRL 73 (1994), 3089.; X.-P. Li et al, PRST-AB 9 (2006), 034201
- [14] W. Henneberg, Annalen der Physik (Leipzig) 19 (1934) 335; W.E. Millett PR 74 (1948) 1058

- [15] A. Noda, M. Ikegami, T. Shirai, New Journal of Physics 8, 288 (2006)
- [16] T. Schätz, U. Schramm, D. Habs, Nature 412, 717 (2001)
- [17] U. Schramm, D. Habs, Progress in Particle and Nuclear Physics 53, 583 (2004).
- [18] Y. Yuri, H. Okamoto, PRST-AB 8 (2005) 114201
- [19] W.D. Phillips, RMP 70 (1998) 721
- [20] V.W. Paul, H. Steinwedel, Zeitschrift für Naturforschung A8, 448 (1953)
- [21] K. Ito et al, IPAC'10, p.3622

## 2.7.4 Ionization Cooling

*D. Neuffer, FNAL*

In ionization cooling ( $\mu$ -cooling), particles pass through a material medium and lose energy (momentum) through ionization interactions, and this is followed by beam reacceleration in rf cavities (Fig.1). The losses are parallel to the particle motion, and therefore include transverse and longitudinal momentum losses; reacceleration restores only longitudinal momentum, and the transverse cooling results from that loss of transverse momentum. The cooling effect is opposed by the random process of multiple scattering in the material medium, which increases the rms beam divergence and therefore the emittance. This cooling method is not generally useful for protons, which would have frequent nuclear interactions, or electrons, which would have bremsstrahlung, but is practical for muons, and cooling rates compatible with muon lifetime are possible.

The differential equation for rms transverse cooling is [1, 2, 3]

$$\begin{aligned} \frac{d\varepsilon_N}{ds} &= -\frac{1}{\beta^2 E} \frac{dE}{ds} \varepsilon_N + \frac{\beta\gamma\beta_\perp}{2} \frac{d\langle\theta_{\text{rms}}^2\rangle}{ds} \\ &= -\frac{g_\perp}{\beta^2 E} \frac{dE}{ds} \varepsilon_N + \frac{\beta_\perp E_s^2}{2\beta^3 m_\mu c^2 L_R E} \end{aligned} \quad (1)$$

where the first term is the energy-loss cooling effect and the second is the multiple-scattering heating term,  $\varepsilon_N$  is the normalized emittance,  $E$  is the beam energy,  $\beta$  and  $\gamma$  are the Lorentz factors,  $dE/ds$  is the energy loss rate,  $\theta_{\text{rms}}$  is the rms multiple scattering angle,  $L_R$  is the material radiation length,  $\beta_\perp$  is the betatron function at the absorber, and  $E_s$  is the characteristic scattering energy ( $\sim 13.6$  MeV). (see also Sec.3.3.1) The symbol  $g_\perp$  is introduced to include coupling with longitudinal cooling (see Eq.(7)); for uncoupled cooling  $g_\perp = 1$ .

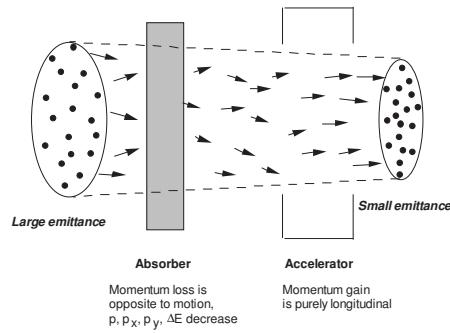


Figure 1: Concept of ionization cooling.

The equation for longitudinal cooling with energy loss is

$$\frac{d\sigma_E^2}{ds} = -2 \frac{\partial \frac{dE}{ds}}{\partial E} \sigma_E^2 + \frac{d\langle\Delta E_{\text{rms}}^2\rangle}{dt} \quad (2)$$

The first term is the cooling term; the second is the heating term caused by random fluctuations in the particle energy loss. Beam cooling can occur if the derivative  $\partial(dE/ds)/\partial E > 0$ . This energy loss can be estimated by the Bethe-Bloch equation (Sec.3.3.1),

$$\begin{aligned} \frac{dE}{ds} &= 4\pi N_A \rho r_e^2 m_e c^2 \frac{Z}{A} \\ &\times \left[ \frac{1}{\beta^2} \ln \left( \frac{2m_e c^2 \gamma^2 \beta^2}{I(Z)} \right) - 1 - \frac{\delta}{2\beta^2} \right] \end{aligned} \quad (3)$$

where  $N_A$  is Avogadro's number,  $\rho$ ,  $A$  and  $Z$  are the density, atomic weight and number of the absorbing material,  $m_e$  and  $r_e$  are the mass and classical radius of the electron,  $(4\pi N_A r_e^2 m_e c^2 = 0.3071 \text{ MeV cm}^2/\text{gm})$ . The ionization constant  $I(Z)$  is approximately  $16Z^{0.9}$  eV, and  $\delta$  is the density effect factor which is small for low-energy muons. The derivative is negative (or naturally heating) for  $E_\mu \lesssim 0.3$  GeV, and is only slightly positive (cooling) for higher energies.

In the long-pathlength Gaussian-distribution limit, the second term in Eq.(2) is approximately

$$d\langle\Delta E_{\text{rms}}^2\rangle = 4\pi (r_e m_e c^2)^2 n_e \gamma^2 \left( 1 - \frac{\beta^2}{2} \right) \quad (4)$$

where  $n_e$  is the electron density in the material. This expression increases rapidly with higher energy (larger  $\gamma$ ), opposing the cooling process. After adding this energy straggling, ionization cooling does not naturally provide adequate longitudinal cooling.

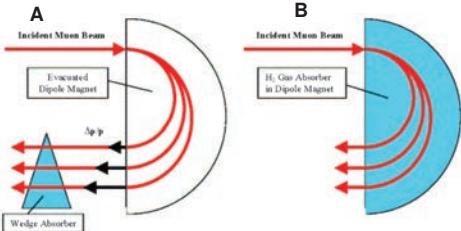


Figure 2: Overview of emittance exchange. A dispersion is introduced into the beam transport. A: Wedge technique: A “wedge” absorber is inserted that presents more path length, and therefore more energy loss, for higher-energy particles is. B: Homogeneous Absorber: The dispersion makes total path length energy dependent; an absorber that fills the transport then obtains energy-dependent energy loss.

However, the cooling term can be enhanced by placing the absorbers where transverse position depends upon energy (nonzero dispersion) and where the absorber density or thickness also depends upon energy, such as in a wedge absorber. This makes the beam particle path length through the absorber material energy dependent (Fig.2). In that case the cooling derivative can be rewritten as

$$\frac{\partial \frac{dE}{ds}}{\partial E} \Rightarrow \frac{\partial \frac{dE}{ds}}{\partial E} \Big|_0 + \frac{dE}{ds} \frac{\eta \rho'}{\beta c p \rho} \quad (5)$$

where  $\rho'/\rho_0$  is the change in density with respect to transverse position,  $\rho_0$  is the reference density associated with  $dE/ds$ , and  $\eta$  is the dispersion. Increasing the longitudinal cooling rate in this manner decreases the transverse cooling by the same amount. The transverse cooling term is changed to

$$\frac{d\varepsilon_N}{ds} = -\frac{1}{\beta^2 E} \frac{dE}{ds} \left( 1 - \frac{\eta \rho'}{\rho_0} \right) \varepsilon_N \quad (6)$$

Note that the coupled transverse cooling (and heating) changes occur in the same direction (i.e. horizontal or vertical) as the dispersion and wedge. However the sum of the cooling rates (over  $x$ ,  $y$ , and  $z$ ) remains constant. This sum can be represented, as with radiation damping, as a sum of cooling partition numbers, defined as the ratios of the cooling rate to the fractional momentum loss rate. For  $x$  and  $y$  emittance cooling the partition numbers are both naturally 1,

$$g_y = g_x = \left( \frac{d\varepsilon_x/ds}{\varepsilon_x} \right) / \left( \frac{dp/ds}{p} \right) = 1 \quad (7)$$

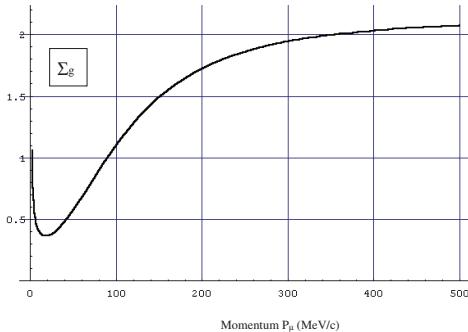


Figure 3: The sum of the cooling partition numbers  $\Sigma_g = (g_x + g_y + g_L)$  as a function of momentum  $P_\mu$  (0-500 MeV/c).  $g_x$  and  $g_y$  are naturally 1 while  $g_L$  becomes strongly negative for  $P_\mu < 200$  MeV/c.  $\Sigma_g$  remains greater than 0, which means that ionization loss remains intrinsically cooling at low momenta.

The longitudinal partition number is

$$g_L = \frac{\frac{d\varepsilon_L/ds}{\varepsilon_L}}{\frac{dp/ds}{p}} = \frac{\partial(\frac{dE}{ds})}{\partial E} \Big/ \frac{dp}{ds} = \frac{\frac{\partial(dE/dt)}{\partial p}}{\frac{(dp/dt)}{p}} \quad (8)$$

and is a function of muon energy. With  $\delta = 0$  in the energy loss formula we find

$$g_L = -\frac{2}{\gamma^2} + \frac{2(1 - \frac{\beta^2}{\gamma^2})}{\left( \ln \left[ \frac{2m_e c^2 \beta^2 \gamma^2}{I(Z)} \right] - \beta^2 \right)} \quad (9)$$

The sum of partition numbers  $\Sigma_g = g_x + g_y + g_L$  is

$$\Sigma_g = 2\beta^2 + 2 \frac{\left( 1 - \frac{\beta^2}{\gamma^2} \right)}{\left( \ln \left[ \frac{2m_e c^2 \beta^2 \gamma^2}{I(Z)} \right] - \beta^2 \right)} \quad (10)$$

This sum is a function of muon momentum, and is displayed in Fig.3 (evaluated at  $H_2$  absorber parameters).  $\Sigma_g$  is approximately 2 for  $P_\mu > 0.3$  GeV/c, but is smaller for lower energies. However,  $\Sigma_g$  does remain positive for all energies, which indicates that cooling remains possible even at low muon energies.

With (horizontal) wedge enhancement of longitudinal cooling,  $g_L$  increases by  $\eta \rho'/\rho_0$ , and  $g_x$  becomes  $g_x = 1 - \eta \rho'/\rho_0$ , leaving the sum of the partition numbers  $\Sigma_g$  constant. (This coupling also mixes the heating terms; in initial approximations we neglect this complication.)

The longitudinal cooling Eq.(2) only tracks energy spread. It can be transformed into a longitudinal emittance cooling equation by adding longitudinal rf focusing that places the beam within

a bunch,

$$\frac{d\varepsilon_L}{ds} = -\frac{g_L}{\beta^2 E} \frac{dE}{ds} \varepsilon_L + \frac{\beta_z}{2} \frac{d\langle \Delta E_{\text{rms}}^2 \rangle}{ds} \quad (11)$$

Here  $\beta_z$  is a focusing function, defined by

$$\beta_z^2 = \frac{\langle \delta z^2 \rangle}{\langle \Delta E^2 \rangle} = \frac{1}{\beta^3 \gamma e V' \sin \phi_s} \frac{\lambda_0}{2\pi} \frac{\alpha_p}{mc^2} \quad (12)$$

where  $\lambda_0$  is the focusing rf wavelength,  $eV' \sin \phi_s$  is the mean focusing rf gradient, and  $\alpha_p$  is the momentum compaction. ( $\alpha_p = 1/\gamma^2 - 1/\gamma_t^2$  in a synchrotron and  $\alpha_p = 1/\gamma^2$  in a linac.) The longitudinal bunch emittance  $\varepsilon_L$  is in  $\delta E - \delta c\tau$  units.

Guidelines for optimum cooling can be obtained from Eqs.(1)-(4) and the partition functions. It is desirable to obtain small  $\beta_\perp$  (from strong focusing) in the absorbers. Strong focusing systems with high-field solenoids or quads or current carrying lenses (Li or Be lens [4], Sec.7.2.12, where the focusing element is also the energy loss medium) are needed. To minimize multiple scattering emittance dilution, materials with large values of the product  $L_R dE/ds$  are preferred, and this is maximal for light element absorbers. Eq.(4) indicates that energy straggling increases greatly with high energies, while the partition function becomes small for low energies (Fig.2). An optimum for cooling with minimal heating occurs at intermediate values; i.e.,  $\gamma \approx 3$  ( $p_\mu \approx 300$  MeV/c or  $B\rho \approx 1$  T-m), where the partition number sum is  $\sim 2$ .

The above formalism assumes uncoupled  $x$  and  $y$  motion, and ignores potentially complicating higher-order effects. Solenoid mixes  $x$  and  $y$  and complicates the motion with angular momentum effects. Wang and Kim have developed rms equations for ionization cooling with quad and solenoid focusing as well as dispersion with wedge focusing, and these include coupling terms in heating and cooling, and angular momentum damping [5]. The additional terms are small in “well-designed” cooling transports (with periodical solenoid field flips and optimized wedge/dispersion insertions), and the 1-D cooling Eqs.(1) and (11) remain good first approximations. In solenoid focusing systems with dispersion/wedge longitudinal cooling, the  $x$  and  $y$  motions can be tightly coupled with approximately equal focusing and cooling; the transverse anti-damping due to the dispersion/wedge is equally split between  $x$  and  $y$ .

Derbenev and Johnson have designed a helical solenoid transport with rf cavities, where tight

coupling of  $x$ ,  $y$ , and  $z$  motion with integrated absorbers enables efficient 3-D cooling [6]. Particles follow helical orbits at a radius  $a$  and period  $\lambda$ , where  $a$  depends on the particle momentum  $p$ , following

$$p(a) = \frac{\sqrt{1+k^2a^2}}{k} \left( eB_s - \frac{eB_h(1+k^2a^2)}{ka} \right)$$

where  $k = 2\pi/\lambda$  is the wave number of the helical period, and  $B_s$  and  $B_h$  are the solenoidal and helical magnetic fields on the reference orbit, respectively. The dispersion  $\eta = p da/dp$  can be calculated from the equation, including the divergence of  $B_s$  and  $B_h$  with respect to radius. By embedding a homogeneous dense hydrogen gas and rf cavities into such a continuous dispersion magnet, longitudinal phase space cooling takes place without a wedge absorber as shown in Fig.2, since the path-length is momentum dependent.

Simulation codes that track the full dynamics of ionization cooling have been developed (ICOOL [7] and G4BeamLine [8]). An international collaboration (MICE) is constructing a cooling section with absorbers, rf and solenoids at RAL and will test it in a muon beam to determine the practicality of ionization cooling [9].

Scenarios that can obtain ionization cooling of muons by the large factors needed for  $\mu^+ - \mu^-$  colliders (Sec.1.6.15) are under investigation [10]. In those studies transverse cooling from  $\varepsilon_N \approx 0.015$  to  $0.000025$  m-rad, and longitudinal cooling by an order of magnitude are required. Scenarios for muon storage ring based neutrino factories use cooling by much smaller factors (i.e., transverse cooling by a factor of  $\sim 10$ ) [11].

It has also been noted that ionization cooling beam dynamics can be useful in low-energy ion and proton storage rings, increasing the lifetime of beams that are stored for multi-turn passages through thin targets, and thus enhancing the nuclear production of useful low-energy secondaries, such as neutrons for medical treatment and unstable nuclei for beta-beams [12].

## References

- [1] A.N. Skrinsky, V.V. Parkhomchuk, Sov. J. Nucl. Physics 12, 3 (1981)
- [2] D. Neuffer, PA 14, 75 (1983)
- [3] D. Neuffer, NIM A 532, 26 (2004)
- [4] B.F. Bayanov et al, NIM 190, 9 (1981)
- [5] C.X. Wang, K.J. Kim, NIM A 532, 260 (2004)
- [6] Y. Derbenev, R.P. Johnson, PRST-AB 8, e041002 (2005)

## Sec.2.7: BEAM COOLING

- [7] R. Fornow, PAC 1999, p.3020,  
<http://pubweb.bnl.gov/people/fornow/>
- [8] T. Roberts et al, (2009),  
<http://g4beamline.muonsinc.com>
- [9] R. Sandstrom, MICE Collab., AIP Conf. Proc. 981:107 (2008)
- [10] R. Palmer et al, PAC 2009
- [11] J.S. Berg et al, PRST-AB 9, 011001 (2006)
- [12] Y. Mori, NIM A562, 591 (2006); C. Rubbia, et al, NIM A568, 475 (2006); D. Neuffer, NIM A585, 109 (2008)

### 2.7.5 Crystalline Beam

*J. Wei, Michigan State U.*

Since mid 1980's, there has been experimental [1, 2] and theoretical [3] efforts to achieve crystalline beams. When the beams in storage rings are sufficiently cold in the beam rest frame, the ions may "lock into" a position where the repelling Coulomb force on the average balances the external focusing force. The interest, besides intrinsically on this new state of matter, is primarily on studying the physics of completely space-charge dominated beams, and the possibility of obtaining high luminosity in colliders [4].

**Conditions of crystallization** There are two necessary conditions [5] to form and maintain a multidimensional crystalline beam: (i) The ring is alternating-gradient (AG) focusing operating below transition, and (ii) The ring lattice periodicity is at least  $2\sqrt{2}$  as high as the maximum betatron tune. Condition (i) arises from the criterion of stable kinematic motion under Coulomb interaction when particles are subject to bending in a storage ring. Condition (ii) arises from the criterion that there is no linear resonance between the phonon modes of the crystalline structure and the machine lattice periodicity. A ring lattice periodicity as high as 4 times the maximum betatron tune is preferred so that linear resonance stopbands are not crossed during the cooling process as the beam density is increased [6]. Condition (ii) is relaxed for the ordering of 1-D states in the azimuthal direction [7, 8].

**Beam rest-frame Hamiltonian [5]** Consider a system of multi-species of ions using the rest frame  $(x, y, z, t)$  of a circulating reference particle. Define the reference particle with electric charge  $Z_0e$  and atomic mass  $M_0$ , and define for the  $i$ th species of ions with charge  $Z_ie$  and mass  $M_i$ ,

$$\bar{Z}_i \equiv Z_i/Z_0, \text{ and } m_i \equiv M_i/M_0 \quad (1)$$

Measure dimensions in units of the characteristic distance  $\xi$  with  $\xi^3 = r_0\rho^2/\beta^2\gamma^2$ , time in units of  $\rho/\beta\gamma c$ , and energy in units of  $\beta^2\gamma^2Z_0^2e^2/4\pi\epsilon_0\xi$ , where  $r_0 = Z_0^2c^2/4\pi\epsilon_0M_0c^2$  is the classical radius,  $\beta c$  and  $\gamma M_0c^2$  are the velocity and energy of the reference particle, and  $\rho$  is the radius of curvature in bending regions of magnetic field  $B_0$ . In a bending region with pure dipole magnetic field, the Hamiltonian for particles of the  $i$ th species is

$$H_i = \frac{1}{2} (P_x^2 + P_y^2 + P_z^2) \quad (2)$$

$$+ \frac{1}{2} \left[ x^2 + \left( \frac{\bar{Z}_i}{m_i} - 1 \right) \gamma^2 z^2 \right] + V_{Ci}$$

$$- \left( 2 - \frac{\bar{Z}_i}{m_i} \right) \gamma x P_z + \left( \frac{\bar{Z}_i}{m_i} - 1 \right) \frac{\rho}{\xi} x$$

where the Coulomb potential is

$$V_{Ci} = \sum_j \frac{\bar{Z}_i \bar{Z}_j / m_i}{\sqrt{(x_j - x)^2 + (y_j - y)^2 + (z_j - z)^2}} \quad (3)$$

The summation,  $j$ , is over all the other particles. In a non-bending region with longitudinal electric field and non-dipole magnetic fields,

$$H_i = \frac{1}{2} (P_x^2 + P_y^2 + P_z^2) - \frac{n_1}{2} \frac{\bar{Z}_i}{m_i} (x^2 - y^2) \quad (4)$$

$$- n_{1s} \frac{\bar{Z}_i}{m_i} xy - \frac{n_2 \xi}{6} \frac{\bar{Z}_i}{m_i} (x^3 - 3xy^2) + V_{Ci} + U_s$$

where the quadrupole, skew quadrupole, and sextupole strengths are represented by  $n_1 = -(\rho/B_0)(\partial B_y/\partial x)$ ,  $n_{1s} = -(\rho/B_0)(\partial B_y/\partial y)$ ,  $n_2 = -(\rho/B_0)(\partial^2 B_y/\partial x^2)$ , respectively, the electrical force  $F_s$  is expressed in terms of electric field  $E_s$  in the laboratory frame,

$$F_s \equiv -\frac{\partial U_s}{\partial z} = \frac{\bar{Z}_i}{m_i} \frac{Z_0 e E_s \xi}{M_0 c^2} \left( \frac{\rho}{\xi \beta \gamma} \right)^2 \quad (5)$$

In the usual case of a single species of ion,  $\bar{Z}_i = m_i = 1$ .

**Numerical methods** Numerical study of the crystalline state has been performed with the molecular dynamics (MD) method [3, 5, 9] iterating the equations of motion derived from the Hamiltonian. Ewald-type summation is performed in the azimuthal direction to evaluate the long-ranged Coulomb forces among particles and their image charges modelled in periodic "super-cells" for computing efficiency.

**Ground-state** In a crystalline ground state, the motion of the circulating particles is periodic in time [5] with the period of the machine lattice.

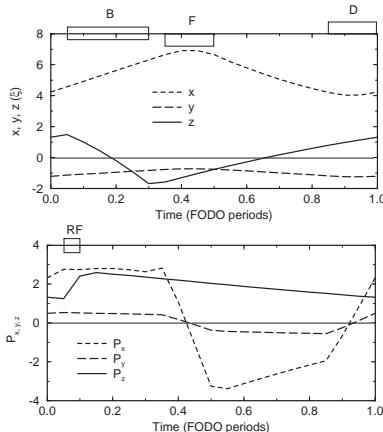


Figure 1: Particle trajectory of a bunched crystalline beam. The machine consists of 10 FODO cells with  $\nu_x = 2.8$ ,  $\nu_y = 2.1$ , and  $\gamma = 1.4$ . Lattice components in each cell are displayed on the figure: B is a bending section, F and D are focusing and de-focusing quadrupoles, and RF is the bunching rf cavity.

As shown in Fig.1, particle trajectory in the transverse direction conforms to AG focusing (breathing), and in the longitudinal direction conforms to the change in bending radius (shear). In the presence of a longitudinal electric field, momentum  $P_z$  also varies periodically conforming to the energy gain at the cavity.

The ground state structure is a 1-D chain when the beam line density is low [5, 9]. The structure becomes 2-D lying in the plane of weaker transverse focusing if the line density  $\lambda$  in the machine is

$$\lambda > 0.62\gamma\xi^{-1} [\min(\nu_y^2, \nu_x^2 - \gamma^2)]^{1/3} \quad (6)$$

where  $\nu_{x,y}$  are the transverse tunes. For even higher density, the particles arrange themselves into 3-D crystals, becoming helices and then helices within helices. Fig.2 shows such a multi-shell structure at a defocusing location of the lattice. The average ratio between the horizontal and vertical particle spacing is about  $[\nu_y^2/(\nu_x^2 - \gamma^2)]^{1/3}$ . The maximum spatial density in the laboratory frame is approximately  $\gamma\nu_y\sqrt{\nu_x^2 - \gamma^2}/(2\xi^3)$ . If a sinusoidal electric field is present, the crystalline structure can be bunched azimuthally.

**Lattice heating** A crystalline beam in its ground state, despite breathing and shear motion, remains in the zero-temperature state [5]. At any

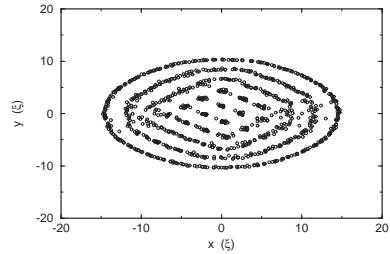


Figure 2: A multi-shell structure with particle positions projected into the  $x$ - $y$  plane ( $\lambda = 25\xi^{-1}$ ).

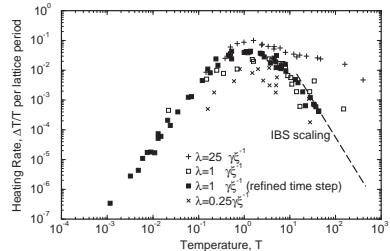


Figure 3: Typical heating rates as functions of temperature obtained by MD simulation at various line densities  $\lambda$ .

nonzero temperature the crystalline beam absorbs energy and heats up under time-dependent external forces caused by variations in lattice focusing and bending. In the high temperature limit, this intrabeam scattering (Sec.2.4.12) results in a growth rate  $\propto \lambda T^{-5/2}$  for coasting beams and  $\propto \lambda T^{-3}$  for bunched beams, where the normalized temperature  $T$  (with its components defined as the deviation of  $P_x$ ,  $P_y$  and  $P_z$  from their ground-state values, squared and averaged over particles) is related to the conventional beam temperature  $T_B$  at high temperature by

$$T \approx \frac{2k_B\rho^2}{\beta^2\gamma^2M_0c^2\xi^2}T_B \quad (7)$$

with  $k_B$  the Boltzmann constant. The peak heating rate occurs at the temperature of about  $T \approx 1$  when the ordering starts to occur, as shown in Fig.3.

Typically, strong spacial correlation appears in all directions when the temperature  $T \lesssim 0.05$ . Lattice shear and AG focusing have similar effects on beam heating. Heating behavior is similar for both bunched and coasting beams. Effects of machine lattice imperfection, ion neutralization, and envelope instability have been studied.

**Cooling methods** To attain a crystalline state, the beam must be effectively cooled in 3-D with a sufficient speed to overcome the heating. Both electron (Sec.2.7.2) and laser (Sec.2.7.3) cooling provide high cooling efficiency in the longitudinal direction, reaching a beam temperature of less than 1 K, but not in the transverse directions (around 100 K) [2]. “Sympathetic cooling” due to intrabeam scattering does produce transverse cooling [10], but the heat exchange becomes ineffective as the beam approaches an ordered state. Coupling cavities operating on a synchrobetatron resonance or regular rf cavities in a dispersive region can provide effective 3-D cooling [11], but the coupling mechanism ceases to work before an ordering can be reached. Realization of multidimensional crystalline beams requires cooling that provides the ions with constant angular velocity, rather than constant linear velocity (so-called tapered cooling) [12].

**Experiments** Experimentally, 3-D ion crystallization has been observed in rf traps [13], static [14] traps, and rf quadrupole rings [15] using laser cooling (Sec.2.7.3). Ion traps were used to experimentally simulate features of an AG-focusing storage ring [16]. “Moving” crystals were generated in a ring-shaped Paul trap system [17].

In storage rings of sizable circumference with beams of significantly high energy, 3-D crystallization has not been successful due to the lack of effective transverse cooling, and the lack of ring lattice superperiodicity. Evidence of anomaly in the Schottky signal was observed on an electron-cooled proton beam at NAPM [1]. 1-D ordering were reported using electron cooling with both proton and heavier ions at storage rings ESR, [18] SIS, CRYRING, and S-LSR. Shear-free ring lattices consisting of both magnets and electrodes were proposed [19] and constructed [20]. Further efforts include developing storage rings of high lattice super-periodicity [21], attempting ultra-low temperature (or emittance) three-dimensionally using laser cooling and transverse-longitudinal resonance coupling, and exploring tapered cooling.

**Applications** The crystalline states correspond to ultralow emittance states of charged-particle beams. Ordered multidimensional beams were proposed for ion-ion colliders for increased luminosity [4]. Machine lattices of high or imaginary transition energy were proposed so that high-energy or colliding crystals may be real-

ized in storage rings of moderate circumference. High-density 1-D strings were proposed for high-luminosity ion-electron collisions with rare radioactive ions [8, 22].

## References

- [1] E.E. Dement'ev et al, *Zh. Tekh. Fiz.* 50 (1980) 1717; N.S. Dikanskii, D.V. Pestrikov, Proc. Workshop on Electron Cooling and Related Applications, KfK 3846 (1984); V.V. Parkhomchuk, A.H. Skrinsky, *Reports on Progress in Physics*, 54 (1991) 919
- [2] S. Schröder et al, *PRL* 64 (1990) 2901; J.S. Hangst et al, *PRL* 67 (1991) 1238
- [3] J.P. Schiffer, P. Kienle, *Z. Phys. A* 321 (1985) 181; A. Rahman, J.P. Schiffer, *PRL* 57 (1986) 1133
- [4] J. Wei, A.M. Sessler, EPAC 98, p.862; J. Wei et al, COOL 07, p.91
- [5] J. Wei, X-P. Li, A.M. Sessler, *PRL* 73 (1994) 3089; BNL-52381 (1993)
- [6] I. Hofmann, J. Struckmeier, Proc. Workshop on Crystalline Ion Beams (1988) p.140; B. Yang et al, *Phys. Plasmas*, 3 (1996) 688; Y. Yuri, H. Okamoto, PRST-AB, 8, 114201 (2005)
- [7] R. Hasse, *PRL* 83 (1999) 3430
- [8] I. Meshkov et al, RIKEN Report: RIKEN- AF-AC-34 (2002)
- [9] R.W. Hasse, J.P. Schiffer, *Ann. Phys.* 203 (1990) 419
- [10] H.-J. Miesner et al, *PRL* 77 (1996) 623
- [11] H. Okamoto, A.M. Sessler, D. Möhl, *PRL* 72 (1994) 3977
- [12] J. Wei, H. Okamoto, A.M. Sessler, *PRL* 80 (1998) 2606; H. Okamoto, J. Wei, *PR E58*, 3817 (1998)
- [13] F. Diedrich et al, *PRL* 59 (1987) 2931; D.J. Wineland et al, *PRL* 59 (1987) 2935
- [14] S.L. Gilbert et al, *PRL* 60 (1988) 2022
- [15] H. Walther, Proc. Workshop on Light Induced Kinetic Effects on Atoms, Ions and Molecules (1991) p.261
- [16] H. Okamoto et al, Proc. 10th Inter. Comp. Accel. Conf. (2009, to be published); S. Ohtsubo et al, PRST-AB 13, 044201 (2010)
- [17] T. Schätz, U. Schramm, D. Habs, *Natur (London)* 412, 717 (2001); U. Schramm, M. Bussmann, D. Habs, *NIM A* 532, 348 (2004)
- [18] M. Steck et al, *PRL* 77, 3803 (1996)
- [19] R. E. Pollock, *Z. Phys. A* 341, 95 (1991)
- [20] M. Ikegami et al, PRST-AB 7, 120101 (2004)
- [21] A. Nada, *NIM A* 532, 150 (2004)
- [22] T. Katayama, D. Möhl, RIKEN Report RIKEN- AF-AC-39 (2002)

# Chapter 3. ELECTROMAGNETIC AND NUCLEAR INTERACTIONS

## 3.1 SYNCHROTRON RADIATION

### 3.1.1 Radiation of a Point Charge

*H. Wiedemann, Stanford U./SSRL*

**Radiation fields** Electromagnetic radiation emitted from relativistic charged particles, especially electrons, as they are accelerated is called synchrotron radiation. Radiation fields, at location  $P$ ,  $\vec{E} = -\frac{\partial}{\partial t}\vec{A} - \nabla\phi$  and  $\vec{B} = \nabla \times \vec{A}$  with retarded Lienard-Wiechert potentials

$$\begin{aligned}\vec{A}(P, t) &= \frac{1}{4\pi\epsilon_0 c} \frac{e}{R} \frac{\vec{\beta}}{1 + \hat{n} \cdot \vec{\beta}} \Big|_{\text{r}} \\ \phi(P, t) &= \frac{1}{4\pi\epsilon_0} \frac{e}{R} \frac{1}{1 + \hat{n} \cdot \vec{\beta}} \Big|_{\text{r}}\end{aligned}\quad (1)$$

where  $\vec{R}(t_r)$  is the distance vector from observer to source,  $\hat{n} = \vec{R}/R$ . All quantities on the r.h.s. are taken at the retarded time  $ct_r = ct - R(t_r)$ .

With  $c\vec{B} = (\vec{E} \times \hat{n})_{\text{r}}$  the Poynting vector becomes  $\vec{S}_{\text{r}}(t) = -\epsilon_0 c^2 E^2 (1 + \vec{\beta} \cdot \hat{n}) \hat{n}|_{\text{r}}$  and the instantaneous differential radiation power is

$$\begin{aligned}\frac{dP(t)}{d\Omega} &= -\hat{n} \cdot \vec{S} R^2|_{\text{r}} = \epsilon_0 c^2 \vec{E}^2 (1 + \vec{\beta} \cdot \hat{n}) R^2|_{\text{r}} \\ &= \frac{r_e m_e c^2}{4\pi c} \frac{R^5}{r^5} \left| \hat{n} \times [(\hat{n} + \vec{\beta}) \times \dot{\vec{\beta}}] \right|^2_{\text{r}}\end{aligned}\quad (2)$$

where  $r = R(1 + \hat{n} \cdot \vec{\beta}) = R(1 - \beta \cos \theta)$  and  $\theta$  the angle between  $\vec{\beta}$  and  $-\hat{n}$ . Replacing  $dW/d\Omega = \int (dP/d\Omega) dt$  and taking the Fourier transform of the electric field results in the spatial and spectral distribution of radiation energy emitted by one electron during a single pass [1, 2],

$$\begin{aligned}\frac{d^2W}{d\omega d\Omega} &= \frac{r_e m_e c^2}{4\pi^2 c} \omega^2 \\ &\times \left| \int_{-\infty}^{\infty} [\hat{n} \times (\hat{n} \times \vec{\beta})] e^{-i\omega(t_r + \frac{R}{c})} dt_r \right|^2_{\text{r}}\end{aligned}$$

Specific radiation characteristics are determined by the particular temporal variation of  $\vec{\beta}$  and  $\hat{n}$  along the particle path.

**Radiation power** The total radiation power per particle in its own system is

$$P^* = \frac{2r_e m_e c^2}{3c} \dot{\beta}^{*2} \quad (3)$$

and in the laboratory system

$$P = \frac{2r_e m_e c^2}{3c} \gamma^2 \left[ \left( \frac{d}{dt} \gamma \vec{\beta} \right)^2 - \left( \frac{d\gamma}{dt} \right)^2 \right] \quad (4)$$

Defining acceleration vectors parallel and orthogonal to the particle motion  $\dot{\vec{\beta}} = \dot{\vec{\beta}}_{||} + \dot{\vec{\beta}}_{\perp}$  splits the total radiation power into parts due to longitudinal and transverse acceleration respectively.

With  $\dot{\vec{\beta}}_{||} = \frac{c}{\gamma^3} \frac{d\gamma}{ds}$  for longitudinal acceleration,

$$P_{||} = \frac{2}{3} c r_e m_e c^2 \left( \frac{d\gamma}{ds} \right)^2 \quad (5)$$

In case of transverse deflection by magnetic field and bending radius  $\rho$  we have  $\dot{\vec{\beta}}_{\perp} = \beta^2 c / \rho$  and

$$P_{\perp}^m = \frac{2}{3} c r_e m_e c^2 \frac{\beta^4 \gamma^4}{\rho^2} \quad (6)$$

and for transverse deflection by electric field  $\mathbf{E}$  the acceleration is  $\dot{\vec{\beta}}_{\perp} = \frac{e}{\gamma m_e c^2} \mathbf{E}$  and

$$P_{\perp}^e = \frac{2}{3} c r_e m_e c^2 \gamma^2 \left( \frac{e \mathbf{E}}{m_e c^2} \right)^2 \quad (7)$$

The classical result is slightly modified by a quantum mechanical correction due to the recoil of the photon on the particle [3],

$$P_{\text{qm}} = P_{\text{classical}} \left( 1 - \frac{55}{16\sqrt{3}} \frac{\varepsilon_c}{E} \right) \quad (8)$$

where  $\varepsilon_c$  is the critical photon energy [Eq.(6), Sec.3.1.3].

### Sec.3.1: SYNCHROTRON RADIATION

Radiation power depends on the mass of the radiating particle like  $1/m^4$ . For protons and electrons of the same total energy,

$$\frac{P_p}{P_e} = \left(\frac{m_e}{m_p}\right)^4 = 8.80 \times 10^{-14}$$

**Relativistic transformations** Transformation of frequency  $\omega$  and unit vectors  $\hat{n}$  from particle frame (\*) (particle moving in  $+z$  direction) to laboratory frame produces the relativistic Doppler effect

$$\omega = \omega^* \gamma (1 + \beta n_z^*) \quad (9)$$

and efficient collimation into the forward direction,

$$n_{x,y} = \frac{n_{x,y}^*}{\gamma(1 + \beta n_z^*)}, \quad n_z = \frac{\beta + n_z^*}{1 + \beta n_z^*} \quad (10)$$

Angles with respect to  $z$ -axis transform with  $n_z = \cos \Theta$  and  $n_x^2 + n_y^2 = \sin^2 \Theta$  etc. like

$$\sin \Theta \approx \frac{\sin \Theta^*}{\gamma(1 + \beta \cos \Theta^*)} \quad (11)$$

Here  $\Theta$  is the emission angle independent of observer. Radiation emitted into all forward space ( $\Theta^* \leq \pi/2$ ) in particle system is collimated to within an angle  $\Theta = \pm 1/\gamma$  in laboratory system.

### References

- [1] J.S.Schwinger, PR 75 (1949) 1912
- [2] H. Wiedemann, Particle Accelerator Physics, 3rd Edition, Springer, Berlin Heidelberg 2007
- [3] J.S. Schwinger, Proc. Nat. Acad. of Sci. USA 40 (1954) 132

#### 3.1.2 Coherent Radiation

H. Wiedemann, Stanford U./SSRL

**Diffraction limit and brightness** Consider a round Gaussian source ( $\sigma_r$ ) and with angular emission distribution ( $\sigma_{r'}$ ). Further, we define a diffraction pattern where a point  $P$  is defined by the coordinates  $\sin \zeta, \psi$ , where  $\zeta$  is the emission angle with respect to the optical axis (Fig.1). The intensity distribution in this pattern is then given by Fraunhofer's diffraction integral [1]

$$I_g(\zeta, \psi) \propto \int_0^\infty e^{-\frac{\rho^2}{2\sigma_r^2}} J_0(k\rho\eta) \rho d\rho \propto \exp\left[-\frac{1}{2}(k\sigma_r\zeta)^2\right] \quad (1)$$

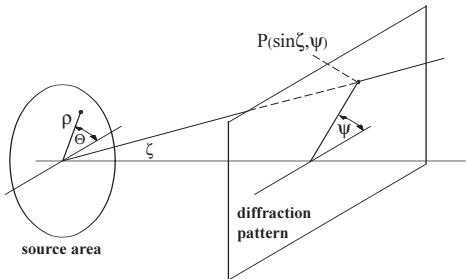


Figure 1: Optical diffraction.

The intensity distribution in  $\zeta$  is still Gaussian and since for small angles  $\zeta \approx \sigma_{r'}$  we find the standard width of the diffraction limited radiation distribution to be  $\sigma_{r'} = 1/(k\sigma_r)$ . The actual geometric intensity distribution in an image plane depends on the distance and particular focusing arrangement.

The diffraction limited photon emittance in the horizontal and vertical plane is wavelength dependent and given by (see also Sec.4.2.2)

$$\epsilon_{ph,u} = \frac{1}{2} \sigma_r \sigma_{r'} = \frac{\lambda}{4\pi}. \quad (2)$$

Radiation from a linear source of length  $L$  (e.g. undulator) appears to emerge from a disk of diameter  $D = L\sigma_{r'}$  in the middle of the source. From diffraction theory it follows for a round aperture that  $D \approx \lambda/\sigma_{r'}$  and therefore with eq.(2)

$$\sigma_{r'} = \sqrt{\frac{\lambda}{L}}, \quad \text{and} \quad \sigma_r = \frac{\sqrt{\lambda L}}{2\pi}. \quad (3)$$

Photon distribution in 6-D phase space is called the photon beam brightness, defined for Gaussian distributions by (Sec.4.2)

$$\mathcal{B} = \frac{\Phi_{ph}}{4\pi^2 \sigma_{Tx} \sigma_{Tx'} \sigma_{Ty} \sigma_{Ty'} (d\omega/\omega)}, \quad (4)$$

where  $\Phi_{ph}$  is the photon flux (1/s) and  $\sigma_{Tu}$ 's are the effective photon source sizes. The diffraction limited brightness is, with  $\sigma_{x,y} = \sigma_r/\sqrt{2}$ , etc.,

$$\mathcal{B}_{max} = \frac{4\Phi_{ph}}{\lambda^2 \frac{d\omega}{\omega}}. \quad (5)$$

For a real particle beam in an undulator, the diffraction limited photon emittance is diluted by finite beam size and divergence. Furthermore, geometrical considerations increase apparent source size in the middle of the undulator, particularly due to (i) path oscillation amplitude  $a = \lambda_p K/(2\pi\gamma)$  in the deflecting plane, where  $\lambda_p$  and

$K$  are the undulator period length and strength (Eq.(6), Sec.3.1.5), respectively; (ii) variation of beam size along the undulator (the  $1/12\dots$ -terms); and (iii) oblique observation at a finite horizontal  $\psi$  and vertical  $\theta$  angle with respect to the optical axis (the  $1/36\dots$ -terms):

$$\begin{aligned}\sigma_{T,x}^2 &= \frac{\sigma_r^2}{2} + \sigma_{b0,x}^2 + a^2 + \frac{\sigma_{b0,x'}^2 L^2}{12} + \frac{\psi^2 L^2}{36} \\ \sigma_{T,x'}^2 &= \frac{1}{2} \sigma_{r'}^2 + \sigma_{b0,x'}^2 \\ \sigma_{T,y}^2 &= \frac{1}{2} \sigma_r^2 + \sigma_{b0,y}^2 + \frac{1}{12} \sigma_{b0,y'}^2 L^2 + \frac{1}{36} \theta^2 L^2 \\ \sigma_{T,y'}^2 &= \frac{1}{2} \sigma_{r'}^2 + \sigma_{b0,y'}^2\end{aligned}\quad (6)$$

where  $\sigma_{b0,u}$  and  $\sigma_{b0,u'}$  are the betatron beam size and divergence at the undulator entrance. To maximize the photon beam brightness the focusing of the electron beam must be adjusted such that 6-D phase space is minimized. For this case  $\sigma_{T,u}$  depends only on  $\sigma_r$  and  $\sigma_{b0,u}$  etc. The optimum matching condition occurs with  $\sigma_{b,u}^2 = \epsilon_u \beta_u$  and  $\sigma_{b0,u'}^2 = \epsilon_u / \beta_u$  when  $\sigma_{T,u} \sigma_{T,u'} = \sqrt{\epsilon_u \beta_x + \frac{1}{2} \sigma_r^2} \sqrt{\epsilon_u / \beta_u + \frac{1}{2} \sigma_{r'}^2}$  becomes a minimum for

$$\beta_u = \frac{\sigma_r}{\sigma_{r'}} = \frac{L}{2\pi}, \quad \text{where } u = x, y \quad (7)$$

### Spatial coherence

$$\frac{\Phi_{coh}}{\Phi_{ph}} = \frac{\lambda^2}{16\pi^2 \sigma_{T,x} \sigma_{T,x'} \sigma_{T,y} \sigma_{T,y'}} \quad (8)$$

defines the fraction of spatially coherent photon flux. Significant spatial coherent radiation is emitted into the forward direction if the beam emittance is reduced to  $\epsilon_u < \lambda/(4\pi)$ .

**Temporal coherence** Particles confined into very small volumes of dimensions  $\lesssim \lambda$  can emit spatial and temporal coherent radiation [2]. The electrical field from electron  $m$  at frequency  $\omega = ck$  has the form  $\mathcal{E}_m \propto e^{i(\omega t + kr_m \cdot \mathbf{n})}$  where  $\mathbf{n}$  the unit vector in the direction from observer to source center and  $\mathbf{r}_m$  is the vector from bunch center to the electron  $m$ . The radiation power from all  $N_e$  electrons is with  $P(\omega) \propto \sum_{m,n=1}^{N_e} \mathcal{E}_m \mathcal{E}_n^*$ ,

$$P(\omega) = p(\omega) N_e [1 + (N_e - 1) g^2(\lambda, \mathbf{r})] \quad (9)$$

where  $p(\omega)$  is the radiation power from a single electron,  $g(\lambda, \mathbf{r}) = |\int e^{ik\mathbf{r}\cdot\mathbf{n}} \Psi(\mathbf{r}) d\mathbf{r}|^2$  the form factor and  $\Psi(\mathbf{r})$  is the three-dimensional, normalized particle distribution. For a uniform

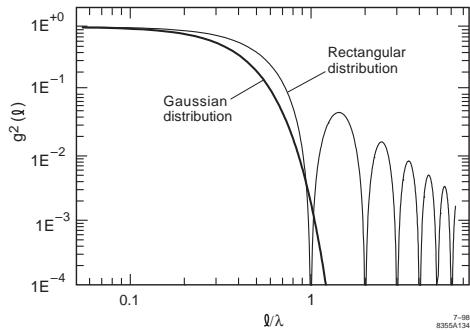


Figure 2: Form factors for Gaussian and uniform particle distribution.

transverse distribution  $\Psi(\mathbf{r}) = \psi(z)$  and either a Gaussian particle distribution  $\psi(z) = \frac{1}{\sqrt{2\pi}\sigma_z} \exp[-z^2/(2\sigma_z^2)]$  with standard width  $\sigma$  or a uniform distribution  $\psi(z) = 1/\ell$  for  $-\frac{1}{2}\ell \leq z \leq \frac{1}{2}\ell$  the form factors become (see Fig.2).

$$g(\lambda, \sigma) = \exp\left(-2\pi^2 \frac{\sigma^2}{\lambda^2}\right) \quad (10)$$

$$g(\lambda, \ell) = \frac{\sin(\pi\ell/\lambda)}{\pi\ell/\lambda} \quad (11)$$

The duration  $\delta t$  and spectral width  $\delta\nu$  of a wavepacket are connected by  $\delta t \delta\nu \geq 1/(4\pi)$ , where  $\delta t$  is called the coherence time and the coherence length is derived from this for the general case and undulator radiation, respectively

$$\ell_{coh} = \frac{\lambda^2}{\Delta\lambda} \quad \text{and} \quad \ell_{coh,und} = N_p \lambda \quad (12)$$

The coherence length can be increased by reducing the spectral width  $\Delta\lambda$  in the monochromator.

### References

- [1] M. Born, E. Wolf, Principles of Optics, Pergamon (1975)
- [2] H. Wiedemann, Particle Accelerator Physics, 3rd Edition, Springer, Berlin, Heidelberg 2007

### 3.1.3 Bending Magnet Radiation

H. Wiedemann, Stanford U./SSRL

Bending radius of charged particle path in homogeneous field  $B$  at energy  $E$  is

$$\frac{1}{\rho} [\text{m}] = \frac{eB}{\beta E} = 0.2998 \frac{B[\text{T}]}{\beta E[\text{GeV}]} \quad (1)$$

### Sec.3.1: SYNCHROTRON RADIATION

Instantaneous radiation power becomes, with  $\beta_{\perp} = \beta^2 c/\rho$ ,

$$P_{\gamma} = \frac{c C_{\gamma}}{2\pi} \frac{E^4}{\rho^2}, \quad \text{with} \quad (2)$$

$$C_{\gamma} = \frac{4\pi}{3} \frac{r_e}{(m_e c^2)^3} = 8.8460 \times 10^{-5} \frac{\text{m}}{\text{GeV}^3}$$

for electrons. The total energy loss to radiation per  $360^\circ$  deflection is

$$U_0 = \oint P_{\gamma} dt = \frac{C_{\gamma}}{2\pi} E^4 \mathcal{I}_2 \quad (3)$$

where  $\mathcal{I}_2$  is defined in Sec.3.1.4. For an isomagnetic ring  $U_0 = C_{\gamma} E^4 / \rho$ . For a circulating beam current  $I$  the total radiation power is

$$P_{\text{tot}}[\text{W}] = 10^9 C_{\gamma} \frac{E^4 [\text{GeV}]}{e\rho} I[\text{A}] \quad (4)$$

The spatial radiation power distribution is ( $\psi$  = deflection angle,  $\theta$  = observation angle with respect to the plane of deflection)

$$\frac{d^2 P_{\gamma}}{d\psi d\theta} = \frac{21}{32} \frac{P_{\gamma}}{2\pi} \gamma \frac{1}{(1 + \gamma^2 \theta^2)^{5/2}} \times \left[ 1 + \frac{5}{7} \frac{\gamma^2 \theta^2}{1 + \gamma^2 \theta^2} \right] \quad (5)$$

The first term represents the  $\sigma$ -mode with the electrical field orthogonal to the deflecting field. The second term is the  $\pi$ -mode for which the electrical field is in the plane of deflecting field and line of observation. After integration,  $P_{\sigma} = (7/8)P_{\gamma}$  and  $P_{\pi} = (1/8)P_{\gamma}$ .

The critical photon energy is

$$\varepsilon_c = \frac{3}{2} \hbar c \gamma^3 / \rho \equiv \hbar \omega_c \quad (6)$$

or (for electrons)

$$\begin{aligned} \varepsilon_c [\text{keV}] &= 2.218 \frac{E[\text{GeV}]^3}{\rho[\text{m}]} \\ &= 0.665 E^2 [\text{GeV}] B[\text{T}] \end{aligned} \quad (7)$$

Spatial spectral photon flux distribution is [1]

$$\frac{d^2 \Phi_{\text{ph}}}{d\theta d\psi} = C_{\Omega} E^2 I \frac{\Delta\omega}{\omega} \frac{\omega^2}{\omega_c^2} K_{2/3}^2(\xi) F(\xi, \theta) \quad (8)$$

where for electrons,

$$\begin{aligned} C_{\Omega} &= \frac{3\alpha}{4\pi^2 e(m_e c^2)^2} \\ &= 1.3255 \times 10^{22} \frac{\text{photons}}{\text{s rad}^2 \text{GeV}^2 \text{A}} \end{aligned} \quad (9)$$

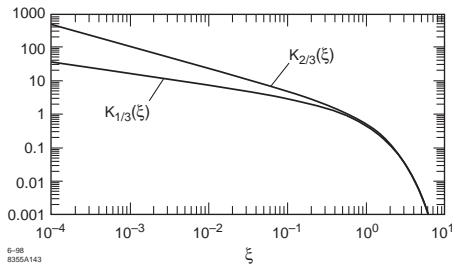


Figure 1: Functions  $K_{1/3}(\xi)$  and  $K_{2/3}(\xi)$ .

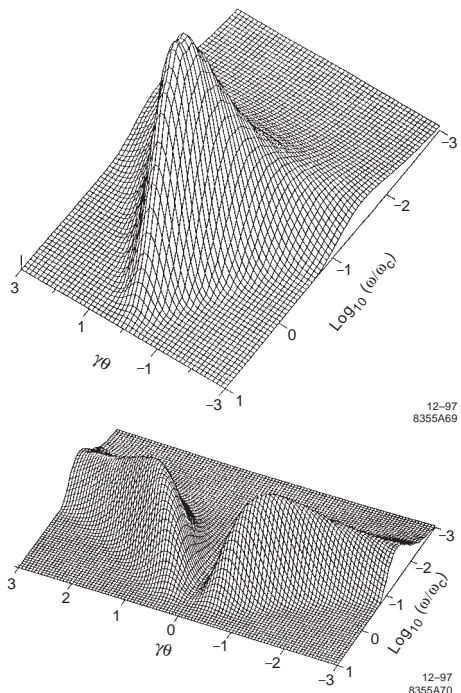


Figure 2: Spatial and spectral distribution of radiation. Upper:  $\sigma$ -mode. Lower:  $\pi$ -mode.

$K_i(\xi)$  is modified Bessel function (Fig.1),  $\xi = \frac{1}{2} \frac{\omega}{\omega_c} (1 + \gamma^2 \theta^2)^{3/2}$ , and

$$F(\xi, \theta) = (1 + \gamma^2 \theta^2)^2 \left[ 1 + \frac{\gamma^2 \theta^2}{1 + \gamma^2 \theta^2} \frac{K_{1/3}^2(\xi)}{K_{2/3}^2(\xi)} \right] \quad (10)$$

Again, the two terms in the square bracket are for the  $\sigma$ - and  $\pi$ - mode polarization respectively. The angular and spectral distributions of both modes are shown in Fig.2, relative intensity

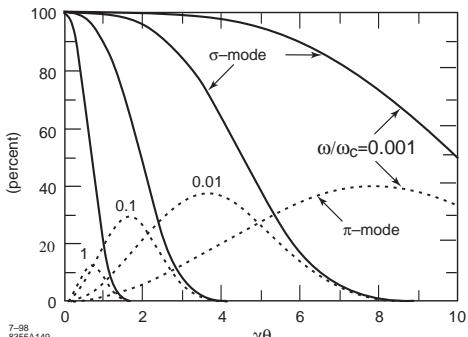


Figure 3: Relative magnitude of  $\sigma$ - and  $\pi$ -mode radiation.

distributions for different photon energies in Fig.3.

Integrating over the nondeflecting plane gives the total photon flux per unit deflection angle,

$$\frac{d\Phi_{ph}}{d\psi} = C_\psi EI \frac{\Delta\omega}{\omega} S\left(\frac{\omega}{\omega_c}\right) \quad (11)$$

with

$$C_\psi = \frac{4\alpha}{9em_ec^2} = 3.9614 \times 10^{19} \frac{\text{photons}}{\text{s rad A GeV}} \quad (12)$$

The universal function (Fig.4) is

$$S(\omega/\omega_c) = \frac{9\sqrt{3}}{8\pi} \frac{\omega}{\omega_c} \int_{\omega/\omega_c}^{\infty} K_{5/3}(x) dx \quad (13)$$

$$\approx \begin{cases} 1.3333 (\omega/\omega_c)^{1/3} & \text{for } \omega \ll \omega_c \\ 0.77736 \frac{\sqrt{x}}{e^x} & \text{for } \omega \gg \omega_c \end{cases}$$

where  $x = \omega/\omega_c$ .

The rms spatial distribution  $\sigma_\theta$  close to the forward direction ( $\theta \approx 0$ ) is, with  $\frac{d\dot{N}_{ph}}{d\psi} \approx \frac{d^2\dot{N}_{ph}}{\theta d\psi} \sqrt{2\pi}\sigma_\theta$ ,

$$\sigma_\theta = \frac{C_\psi}{\sqrt{2\pi}C_\Omega} \frac{S(x)}{x^2 K_{2/3}^2(x/2)} \frac{1}{E} = \frac{f(x)}{E} \quad (14)$$

$f(x)$  is shown in Fig.5.

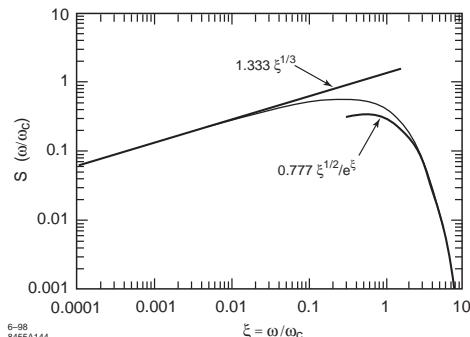


Figure 4: Universal Function  $S(\omega/\omega_c)$ .

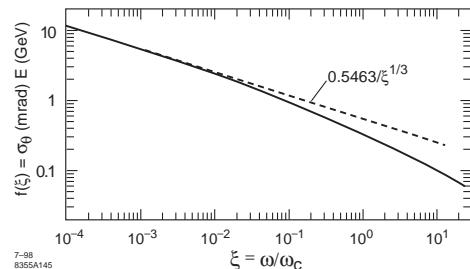


Figure 5: Function  $f(x) = \sigma_\theta E$ .

**Harmonic representation** Synchrotron radiation is emitted at frequencies which are  $\nu$ th harmonics of the revolution frequency up to very high orders determined by the critical photon frequency at harmonic number  $\nu_c = \omega_c/\omega_L = (3/2)\gamma^3$  where  $\omega_L = c/\rho$  is the Larmor frequency. For  $\nu_c \gg 1$  the modified Bessel functions  $K_{1/3}$  and  $K_{2/3}$  in the general expression (8) can be approximated by ordinary Bessel functions of high order  $\nu$  [3, 4] and the photon flux from a circulating beam current  $I$  becomes

$$\frac{d\dot{N}_{ph}}{d\Omega} = \alpha\nu^2 \frac{I}{e} (J_\nu'^2 + \theta^2 J_\nu^2) \frac{\Delta\nu}{\nu} \quad (15)$$

where the arguments of the Bessel functions are  $\nu\beta \cos\theta$ .

**Shielding** Radiation formulas generally hold only in free space. When radiation is emitted within metallic pipes, long wavelength radiation with  $\lambda \gtrsim$  pipe dimensions is cut off. For the case of a beam between two infinite metallic plates at a

### Sec.3.1: SYNCHROTRON RADIATION

distance  $d$ , this shielding effect gives a differential photon flux in the forward direction ( $\theta = 0$ ) [2],

$$\frac{d\dot{N}_{\text{ph}}}{d\Omega} = \alpha\nu^2 \left( \frac{2\pi\rho}{d\nu} \right) \frac{\Delta\nu}{\nu} \frac{I}{e} \quad (16)$$

$$\times \sum_{j=1,3,5,\dots}^{j \leq \nu d \beta / \pi \rho} \left[ J_{\nu}^{\prime 2}(x) + \frac{j\pi\rho}{d} \frac{J_{\nu}^2(x)}{x^2} \right]$$

where  $x = \sqrt{(\nu\beta)^2 + (j\pi\rho/d)^2}$  and  $\nu \gg 1$ . For  $d \rightarrow \infty$ , it becomes Eq.(15).

Shielding is particularly significant for short bunches at frequencies  $\omega \ll \omega_c$ . For a uniform particle distribution of length  $\ell$ , the total coherent radiation power for circulating beam current  $I$  in absence of shielding is

$$P_{\text{coh}}^o = \frac{\pi}{\epsilon_0 c} \left( \frac{\sqrt{3}\rho}{\ell} \right)^{4/3} I^2 \quad (17)$$

and with shielding by two infinite plates at distance  $d$ ,

$$P_{\text{coh}}^{\infty} = \frac{\pi}{\epsilon_0 c} \frac{\sqrt{3}\rho d}{2\ell^2} I^2 \quad (18)$$

The shielding efficiency is  $P_{\text{coh}}^{\infty}/P_{\text{coh}}^o = \sqrt{3}d/(23^{2/3}\ell^{2/3}\rho^{1/3})$ .

## References

- [1] H. Wiedemann, Particle Accelerator Physics II, Springer, Berlin Heidelberg 1995.
- [2] J.S. Nodvick, D.S. Saxon, PR 96 (1954) 180
- [3] A.A. Sokolov, I.M. Ternov, Synchrotron Radiation, Pergamon, Oxford 1968
- [4] G.A. Schott, Phil. Mag. 13 (1907) 194

### 3.1.4 Synchrotron Radiation in Storage Rings

*H.Wiedemann, Stanford U./SSRL*

#### 3.1.4.1 Radiation integrals

Beam parameters in a storage ring are modified by the emission process of synchrotron radiation [1, 2]. See also Sec.2.1.4. Far from linear coupling resonances, they can be expressed approximately through the following radiation integrals [3], where we use the coordinate system and sign convention of Sec.2.1.1 and with  $u = x$  or  $y$  (note

also the numbering scheme in [4]):

$$\mathcal{I}_1[\text{m}] = \oint (\kappa_x D_x + \kappa_y D_y) dz \quad (1)$$

$$\mathcal{I}_2[\text{m}^{-1}] = \oint \kappa^2 dz \quad (2)$$

$$\mathcal{I}_3[\text{m}^{-2}] = \oint |\kappa|^3 dz \quad (3)$$

$$\mathcal{I}_{4x}[\text{m}^{-1}] = \oint [\kappa^2 \kappa_x D_x + 2\kappa_x (kD_x + \underline{k}D_y)] dz \quad (4)$$

$$\mathcal{I}_{4y}[\text{m}^{-1}] = \oint [\kappa^2 \kappa_y D_y + 2\kappa_y (kD_x - \underline{k}D_y)] dz \quad (5)$$

$$\mathcal{I}_{5u}[\text{m}^{-1}] = \oint |\kappa|^3 \mathcal{H}_u dz \quad (6)$$

$$\mathcal{I}_{6x}[\text{m}^{-1}] = \oint (kD_x + \underline{k}D_y)^2 dz \quad (7)$$

$$\mathcal{I}_{6y}[\text{m}^{-1}] = \oint (\underline{k}D_x - kD_y)^2 dz \quad (8)$$

where  $k, \underline{k}$  are the strengths for normal and skew quadrupoles, respectively,  $u = x$  or  $y$ ,  $D_u$  is the dispersion,  $\mathcal{I}_{4u}$  is for a sector magnet,  $\mathcal{H}_u(z) = \beta_u D_u'^2 + 2\alpha_u D_u D_u' + \gamma_u D_u^2$ ,  $\kappa_x \equiv 1/\rho_x$  denotes the horizontal curvature,  $\kappa_y \equiv 1/\rho_y$  the vertical one, and  $\kappa^2 = \kappa_x^2 + \kappa_y^2$ . The above integral expressions include vertical bending, but are missing terms related to coupled normal mode dispersions (Sec.4.7.4.3) [5].

For a wedge magnet the integrals  $\mathcal{I}_{4u}$  must be modified to

$$\mathcal{I}_{4u}|_{\text{wedge}} = \kappa_u^2 D_{u0} \tan \theta_0 + \mathcal{I}_{4u} + \kappa_u^2 D_{ue} \tan \theta_e \quad (9)$$

where  $D_{u0}, D_{ue}$  the dispersion functions at the entrance and exit, respectively and  $\theta_0, \theta_e$  are the entrance and exit angles as measured from a pole face normal to the beam orbit and the signs are chosen such that for a symmetric rectangular magnet  $\theta_0$  and  $\theta_e$  are negative.

#### 3.1.4.2 Radiation damping

Radiation damping occurs due to average energy loss into synchrotron radiation. Damping for energy deviation and bunch length derives from energy dependence of radiation. In the transverse planes damping comes from the loss of transverse momentum with the emission of photons while rf-system replenishes only the longitudinal momentum.

**Damping decrements** Betatron and synchrotron oscillation amplitudes are damped like  $A_i = A_{i,0} e^{-\alpha_i t}$ , where  $i = x, y, z$ . With the circumference  $C$  and

$$\langle P_\gamma \rangle = \frac{1}{C} \oint P_\gamma dz = \frac{cC_\gamma}{2\pi C} E^4 \mathcal{I}_2 \quad (10)$$

where  $C_\gamma$  is defined in eq.(3.1.3) (2) and for a flat ring of circumference  $C$  the damping decrements [6] are

$$\begin{aligned} \alpha_x &= \frac{C_\alpha}{C} E^3 \mathcal{I}_2 \left( 1 - \frac{\mathcal{I}_{4x}}{\mathcal{I}_2} \right) \\ \alpha_y &= \frac{C_\alpha}{C} E^3 \mathcal{I}_2 \left( 1 - \frac{\mathcal{I}_{4y}}{\mathcal{I}_2} \right) \\ \alpha_z &= \frac{C_\alpha}{C} E^3 \mathcal{I}_2 \left( 2 + \frac{\mathcal{I}_{4x} + \mathcal{I}_{4y}}{\mathcal{I}_2} \right) \end{aligned} \quad (11)$$

where  $C_\alpha = \frac{cr_c}{3(mc^2)^3} = 2113.1 \text{ m}^2/\text{GeV}^3/\text{sec.}$

If vertical bending magnets are present then the vertical damping decrement must be modified accordingly similar to the horizontal decrement.

Damping partition numbers  $J_i$  are defined,

$$\begin{aligned} J_x &= 1 - \frac{\mathcal{I}_{4x}}{\mathcal{I}_2}, \quad J_y = 1 - \frac{\mathcal{I}_{4y}}{\mathcal{I}_2} \\ J_z &= 2 + \frac{\mathcal{I}_{4x} + \mathcal{I}_{4y}}{\mathcal{I}_2}, \end{aligned} \quad (12)$$

with Robinson's sum rule [7, 8]

$$\sum_i J_i = 4 \quad (13)$$

The damping time  $\tau_i$  is the reciprocal of the damping decrement  $\alpha_i$ .

**Dependence on rf frequency** The electron beam energy in a storage ring can be changed by varying the rf-frequency  $\Delta\delta \equiv \frac{\Delta p}{p_0} = -\frac{1}{\eta_c} \frac{\Delta f_{rf}}{f_{rf}}$ , where  $\eta_c = \alpha_c - \frac{1}{\gamma^2}$  is the slip-page factor. When displacing the particle orbit transversely, quadrupoles also act like bending magnets with  $\kappa_x = (kD_x + \underline{k}D_y)\delta$  and  $\kappa_y = (-\underline{k}D_x + kD_y)\delta$ . The change in  $\mathcal{I}_{4u}$  is [9] with  $u = x, y$

$$\Delta\mathcal{I}_{4u} = 2\mathcal{I}_{6u}\delta \quad (14)$$

**Dependence on synchrotron oscillation** Oscillating particle energy  $\delta = \delta_{\max} \sin \omega_s t$  due to synchrotron oscillations causes a periodic variation of damping decrements by [9]

$$\frac{\Delta\alpha_z}{\alpha_z} = -\frac{\Delta\alpha_u}{\alpha_u} = \frac{\mathcal{I}_{6u}}{\mathcal{I}_2} \delta_{\max} \sin \omega_s t \quad (15)$$

where  $u = x, y$ . These oscillations result in a "breathing" variation of betatron (and synchrotron oscillation) amplitudes  $a_u$  by

$$\frac{\Delta a_u}{a_{0,u}} = 4 \frac{\mathcal{I}_{6u} \delta_{\max}}{\tau_u \omega_s \mathcal{I}_2} \quad (16)$$

### 3.1.4.3 Quantum excitation

Emission of photons into synchrotron radiation occurs statistically leading to quantum fluctuation of beam parameters [6]. The effect of quantized emission of photons depends on the following quantities:

$$\begin{aligned} \text{mean photon energy } \langle \varepsilon \rangle &= \frac{8}{15\sqrt{3}} \epsilon_{\text{cr}} \\ \text{rms photon energy } \langle \varepsilon^2 \rangle &= \frac{11}{27} \epsilon_{\text{cr}}^2 \\ \text{total photon flux } \dot{N}_{\text{ph}} &= \frac{15\sqrt{3}}{8} \frac{P_\gamma}{\epsilon_{\text{cr}}} \\ \text{product } \dot{N}_{\text{ph}} \langle \varepsilon^2 \rangle &= \frac{55}{24\sqrt{3}} \epsilon_{\text{cr}} P_\gamma \\ \text{product } &= \frac{55}{24\sqrt{3}} \hbar c^2 r_e m_e c^2 |\kappa|^3 \gamma^7 \end{aligned}$$

Here  $\epsilon_{\text{cr}}$  is the critical photon energy [Eq.(6), Sec.3.1.3]. Quantum excitation of beam energy spread along path of length  $L$  is with the fine structure constant  $\alpha$

$$\Delta\sigma_E^2|_q = \frac{55\alpha(\hbar c)^2}{48\sqrt{3}} \gamma^7 \int_0^L |\kappa|^3 dz. \quad (17)$$

Statistical emission of photons causes sudden change in particle orbit at locations where dispersion is finite. This effect leads to statistical increase of betatron oscillations and emittance. Increase of beam emittance  $\epsilon_u$  ( $u = x, y$ ) due to quantum excitation along distance  $L$  is

$$\Delta\epsilon_u = \frac{55 r_e \hbar c}{24\sqrt{3} mc^2} \gamma^5 \int_0^L |\kappa_u^3| \mathcal{H}_{uds} ds. \quad (18)$$

### 3.1.4.4 Equilibrium beam emittances

**Energy spread and bunch length** Equilibrium energy spread is reached when the average  $\langle \rangle_z$  quantum excitation rate around the ring  $\langle d\sigma_E^2/dt|_q \rangle_z$  is equal to damping rate  $\langle d\sigma_E^2/dt|_d \rangle_z = -2\alpha_z \sigma_E^2$ ,

$$\frac{\sigma_E^2}{E^2} = C_q \gamma^2 \frac{\mathcal{I}_3}{2\mathcal{I}_2 + \mathcal{I}_{4x} + \mathcal{I}_{4y}} \quad (19)$$

with

$$C_q = \frac{55}{32\sqrt{3}} \frac{\hbar c}{mc^2} = 3.8319 \times 10^{-13} \text{ m} \quad (20)$$

for electrons. The bunch length is

$$\sigma_z = \frac{c|\eta_c| \sigma_E}{\omega_s E_0} = \frac{\sqrt{2\pi} c}{\omega_0} \sqrt{\frac{\alpha_c E_0}{heV_{rf} \cos \phi_s} \frac{\sigma_E}{E_0}}, \quad (21)$$

where  $\phi_s$  is the synchronous phase determined by  $V_{rf} \sin \phi_s = U_0$  with the sign convention  $\eta \cos \phi_s < 0$  and  $U_0$  given by Sec.3.1.3,  $\omega_s$  is the synchrotron oscillation frequency and the slip-page factor  $\eta_c = \alpha_c - 1/\gamma^2$ .

### Sec.3.1: SYNCHROTRON RADIATION

**Transverse beam emittance** Equilibrium transverse emittance is reached when quantum excitation is equal to damping  $\langle d\epsilon_u/dt|_d \rangle_z = -2\alpha_u \epsilon_u$ :

$$\epsilon_u = C_q \frac{\gamma^2}{J_u} \frac{\mathcal{I}_5}{\mathcal{I}_2} \quad (22)$$

where  $u = x$  or  $y$ . In an ideal, flat accelerator  $D_y = D'_y = 0$  and therefore  $\epsilon_y$  becomes zero. In this case, a transverse recoil due to photon emission into an angle  $1/\gamma$  cannot be ignored anymore and defines a lower fundamental limit on beam emittance,

$$\epsilon_y = \frac{C_q \langle \beta_y \rangle_z}{2J_y} \frac{\mathcal{I}_3}{\mathcal{I}_2} \quad (23)$$

Due to construction and alignment imperfections both horizontal and vertical betatron oscillations can become coupled. The degree of emittance coupling depends on the tunes and the coupling coefficient which is for linear coupling by quadrupoles

$$\kappa_- = \left| \frac{1}{2\pi} \int_s^{s+L} \underline{k} \sqrt{\beta_x \beta_y} \exp(i\Psi) dz \right| \quad (24)$$

where  $\Psi = \psi_x + \psi_y - (\nu_x + \nu_y + q) 2\pi z/L$  with  $\psi_u(z)$  the betatron phases,  $\nu_u$  the tunes,  $m, q$  integers and  $\underline{k}$  the strengths of the skew quadrupoles. In case of a linear sum resonance, both emittances grow indefinitely if  $|\nu_x + \nu_y - p| \leq \kappa$  with  $p$  being an integer. In case of a difference resonance,  $\nu_x - \nu_y$  is close to an integer and we observe a constant exchange of horizontal and vertical beam emittances at a frequency  $\Omega = \sqrt{\kappa^2 + \Delta^2}/2$  where  $\Delta = \nu_x - \nu_y - p$ . The maximum beam emittances obey in accordance with Liouville's theorem the sum rule

$$\epsilon_x + \epsilon_y = \epsilon_{x0} \quad (25)$$

where  $\epsilon_{x0}$  is the uncoupled horizontal emittance. The ratio of the beam emittances is

$$\frac{\epsilon_y}{\epsilon_x} = \frac{\kappa_-^2/2}{\kappa_-^2/2 + \Delta^2} \quad (26)$$

#### 3.1.4.5 Damping wigglers

Wiggler magnets cause additional damping and quantum excitation modifying the equilibrium beam emittance to [9]

$$\begin{aligned} \epsilon_w &= C_q \frac{\gamma^2}{J_u} \frac{\mathcal{I}_5^0 + \mathcal{I}_5^w}{\mathcal{I}_2^0 + \mathcal{I}_2^w}, \quad \text{or} \\ \frac{\epsilon_w}{\epsilon_0} &= \frac{1 + \mathcal{I}_5^w/\mathcal{I}_5^0}{1 + \mathcal{I}_2^w/\mathcal{I}_2^0} \end{aligned} \quad (27)$$

where the indices <sup>0</sup> and <sup>w</sup> indicate that the integral be taken over regular ring bending magnets and wiggler magnets separately. A horizontally deflecting wiggler magnet with  $N_p$  periods and sinusoidal field variation along the electron path in a dispersion-free section of a circular accelerator generates a new equilibrium emittance

$$\frac{\epsilon_{xw}}{\epsilon_{x0}} = \frac{1 + \frac{8C_q}{30\pi J_x} N_p \frac{\beta_x}{\epsilon_{x0} \rho_w} \gamma^2 \frac{\rho_0}{\rho_w} \Theta_w^3}{1 + \frac{1}{2} N_p \frac{\rho_0}{\rho_w} \Theta_w} \quad (28)$$

where  $\Theta_w = \lambda_p/(2\pi\rho_w)$ ,  $\rho_0 > 0$ , the bending radius  $\rho_w > 0$  is determined by the wiggler peak field. The resulting beam emittance is reduced if  $\frac{8}{15\pi} \frac{C_q}{J_x} \frac{\beta_x}{\epsilon_{x0} \rho_w} \gamma^2 \Theta_w^2 \leq 1$ . For very long wigglers ( $N_p \rightarrow \infty$ ) the emittance approaches

$$\frac{\epsilon_{xw}}{\epsilon_{x0}} \rightarrow \frac{8}{15\pi} \frac{C_q}{J_x} \frac{\beta_x}{\epsilon_0 \rho_w} \gamma^2 \Theta_w^2 \quad (29)$$

The change of equilibrium energy spread due to damping wigglers is

$$\frac{\sigma_{ew}^2}{\sigma_{e0}^2} = \frac{1 + \mathcal{I}_3^w/\mathcal{I}_3^0}{1 + \mathcal{I}_2^w/\mathcal{I}_2^0} \quad (30)$$

which approaches (for  $N_p \rightarrow \infty$ )

$$\frac{\sigma_{ew}^2}{\sigma_{e0}^2} \rightarrow \frac{\rho_0}{\rho_w} = \frac{B_w}{B_0} \quad (31)$$

#### 3.1.4.6 Quantum lifetimes

Particle losses occur for Gaussian particle distributions in 6-D phase space due to finite transverse apertures or energy acceptance. Lifetimes due to these effects are [10, 11]

$$\tau_q = \frac{1}{2} \tau_u \frac{e^\xi}{\xi}, \quad \text{with} \quad \xi = \frac{A_u^2}{2\sigma_u^2} \quad (32)$$

where  $u = x, y$  or  $E$  and  $A_u$  the limiting half apertures. For the synchrotron dimension,  $A_s$  might be determined by the rf-bucket height (see also Sec.2.1.2),

$$\begin{aligned} \xi &= \frac{\epsilon_{max}^2}{2\sigma_E^2}, \quad \left( \frac{\epsilon_{max}}{E_0} \right)^2 = \frac{U_0}{\pi |\eta| h E_0} F(q) \\ \text{with } F(q) &= 2 \left[ \sqrt{q^2 - 1} - \cos^{-1}(1/q) \right] \end{aligned} \quad (33)$$

and  $q = eV_{rf}/U_0$  is the overvoltage factor.

## References

- [1] A. Chao, JAP 50 (1979) 595
- [2] K. Hirata, K. Ohmi, K. Oide, PR E49 (1994) 751
- [3] R.H. Helm et al, PAC 73, San Francisco, 1973
- [4] J. Jowett, AIP Proc. 153 (1985) p.934
- [5] D. Sagan, BMAD manual vers.16.11, Sec.10.15, <https://www.lepp.cornell.edu/~dcs/bmad>
- [6] M. Sands, in: Physics with Intersecting Storage Rings, ed. B.Touschek, Academic Press, NY, 1971
- [7] K.W. Robinson, PR 111 (1958) 373
- [8] Iu.F. Orlov, E.K. Tarasov, JETP 34 (1958) 651
- [9] H. Wiedemann, Particle Accelerator Physics, 3rd. Edition, Springer, Berlin, Heidelberg 2007
- [10] A.W. Chao, Lecture Notes Phys., V.296, Springer, Berlin, 1988, p.51
- [11] Y.H. Chin, DESY 87-062 (1987)

### 3.1.5 Undulator and Wigglers Radiation

H. Wiedemann, Stanford U./SSRL

Field expansion for sinusoidal field variation  $B_y = B_0 \cos(k_p z)$  along axis [1]:

$$B_x = 0$$

$$B_y = B_0 \cosh k_p y \cos k_p z \quad (1)$$

$$B_z = -B_0 \sinh k_p y \sin k_p z$$

where  $\lambda_p = 2\pi/k_p$  is the period length. Undulator field strength for hybrid magnet design [2] is, for small gap apertures  $g < \lambda_p$ ,

$$B_y \approx 3.33 \exp \left[ -\frac{g}{\lambda_p} \left( 5.47 - 1.8 \frac{g}{\lambda_p} \right) \right] \quad (2)$$

Equation of motion  $\frac{\dot{v}}{\rho} = \frac{e}{m_e \gamma v_0^2} \vec{v} \times \vec{B}$  has solutions

$$\begin{aligned} x(t) &= \frac{K}{\gamma k_p} \cos(k_p \bar{\beta} c t) \\ z(t) &= \bar{\beta} c t + \frac{K^2}{8k_p \gamma^2} \sin(2k_p \bar{\beta} c t) \end{aligned} \quad (3)$$

with average drift velocity

$$\bar{\beta} = \beta \left( 1 - \frac{K^2}{4\gamma^2} \right), \quad (4)$$

maximum oscillation amplitude  $a$  and deflection angle  $\vartheta$  with respect to axis

$$a = \frac{K}{\gamma k_p} \quad \text{and} \quad \vartheta = \frac{B_0}{B_\rho} \frac{\lambda_p}{2\pi} \quad (5)$$

and strength parameter

$$K = \frac{eB_0}{\beta m_e c k_p} = \gamma \vartheta = 93.4 B \lambda_p \quad (6)$$

A periodic magnetic field device is called an undulator if  $K \lesssim 1$  and a wiggler magnet if  $K \gg 1$ . Radiation is emitted in a line spectrum at a fundamental wavelength  $\lambda_1$  and its harmonics  $i$ . The wavelength [3] is the result of Lorentz contraction and Doppler shift of the magnetic period  $\lambda_p$ ,

$$\lambda_i = \frac{\lambda_p}{2\gamma^2 i} [1 + \frac{1}{2} K^2 + \gamma^2 (\theta^2 + \psi^2)] \quad (7)$$

where  $\theta, \psi$  are the horizontal and vertical observation angles with respect to the axis. The wavelengths for  $\theta = \psi = 0$  are in practical units for electrons

$$\lambda_i [\text{\AA}] = 1305.6 \frac{\lambda_p}{i E^2} (1 + \frac{1}{2} K^2) \quad (8)$$

and the corresponding photon energies

$$\epsilon_i [\text{eV}] = 9.4963 \frac{i E^2}{\lambda_p (1 + \frac{1}{2} K^2)} \quad (9)$$

The spectral bandwidth of undulator radiation is

$$\frac{\Delta \lambda}{\lambda} = \frac{1}{i N_p} \quad (10)$$

and the polar opening angle

$$\sigma_{u'} \approx \frac{1}{\gamma} \sqrt{\frac{1 + \frac{1}{2} K^2}{2i N_p}}, \quad (11)$$

where  $u = x$  or  $y$ . The total energy loss of an electron passing through an  $N_p$ -period wiggler or undulator magnet of length  $L_u = N_p \lambda_p$  is

$$E_{\text{rad}} = \frac{1}{3} r_e m_e c^2 \gamma^2 K^2 k_p^2 L_u. \quad (12)$$

$$E_{\text{rad}} [\text{eV}] = 0.07257 \frac{E [\text{GeV}]^2 K^2}{\lambda_p [\text{m}]^2} L_u [\text{m}]$$

A circulating beam current  $I$  generates a total photon beam power of

$$P = \frac{4\pi^2 r_e}{3 em_e c^2} \frac{E^2 K^2 N_p I}{\lambda_p}, \quad (13)$$

$$P [\text{W}] = 0.07257 \frac{E [\text{GeV}]^2 K^2 N_p I [\text{A}]}{\lambda_p [\text{m}]}. \quad (14)$$

Spectral and spatial photon flux from beam current  $I$  into harmonic  $i$  and  $\sigma$ - and  $\pi$ -mode polarization (Fig.1)

$$\begin{aligned} \frac{d\Phi_{\text{ph}}(\omega)}{d\Omega} &= \alpha \gamma^2 N_p^2 \frac{\Delta \omega}{\omega} \frac{I}{e} \\ &\times \sum_{i=1}^{\infty} i^2 \text{Sinc}(F_{\sigma}^2 + F_{\pi}^2) \end{aligned} \quad (14)$$

### Sec.3.1: SYNCHROTRON RADIATION

with

$$\text{Sinc} = \left( \frac{\sin \pi N_p \Delta \omega_i / \omega_1}{\pi N_p \Delta \omega_i / \omega_1} \right)^2,$$

$$F_\sigma = \frac{2\gamma\theta\Sigma_1 \cos \varphi - K\Sigma_2}{1 + \frac{1}{2}K^2 + \gamma^2\theta^2},$$

$$F_\pi = \frac{2\gamma\theta\Sigma_1 \sin \varphi}{1 + \frac{1}{2}K^2 + \gamma^2\theta^2},$$

where  $\Delta\omega_i = \omega - \omega_i$ , and

$$\Sigma_1(i) = \sum_{m=-\infty}^{\infty} J_{-m}(u) J_{i-2m}(v),$$

$$\Sigma_2(i) = \sum_{m=-\infty}^{\infty} J_{-m}(u) [J_{i-2m-1}(v)$$

$$+ J_{i-2m+1}(v)],$$

$$u = \frac{\omega}{\omega_1} \frac{\bar{\beta} K^2}{4(1 + \frac{1}{2}K^2 + \gamma^2\theta^2)},$$

$$v = \frac{\omega}{\omega_1} \frac{2\bar{\beta} K^2 \gamma \theta \cos \varphi}{1 + \frac{1}{2}K^2 + \gamma^2\theta^2},$$

and the angle  $\theta$  is taken with respect to the beam axis and  $\varphi$  is the azimuthal angle counted counter-clockwise while viewing the source against beam direction with  $\varphi = 0$  in the deflection plane.

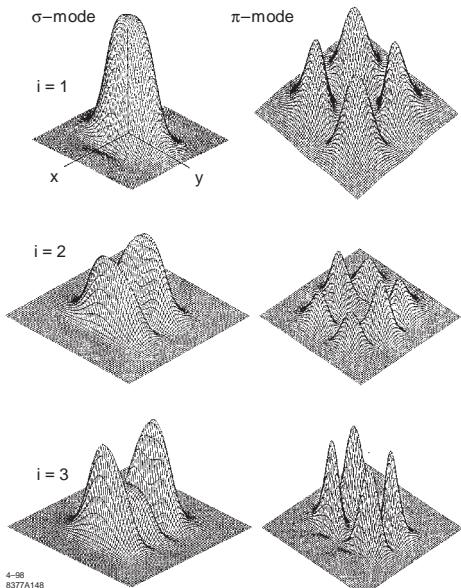


Figure 1: Spatial undulator radiation distribution for  $\sigma$ - and  $\pi$ - mode radiation in lowest order harmonics.

In forward direction ( $\theta = 0$ ) and at harmonic frequencies ( $\Delta\omega_i = 0$ ) the spatial photon flux distribution for the  $i$ -th harmonic is

$$\frac{d\Phi_{\text{ph}}(\omega)}{d\Omega} \Big|_i = \alpha \gamma^2 N_p^2 \frac{\Delta\omega}{\omega} \frac{I}{e} \frac{K^2 i^2 [JJ]}{(1 + \frac{1}{2}K^2)^2}$$

$$= 1.7466 \times 10^{23} E[\text{GeV}]^2 I[\text{A}] N_p^2 \frac{\Delta\omega}{\omega} f_i(K)$$

with

$$[JJ] = \left[ J_{\frac{i-1}{2}}(x) - J_{\frac{i+1}{2}}(x) \right]^2 \quad (15)$$

and  $x = \frac{iK^2}{4+2K^2}$ . The functions  $f_i(K)$  are shown in Fig.2.

Table 1: Tabulation of functions  $f_i(K)$ .

| $K$  | $f_1$ | $f_3$ | $f_5$ | $f_7$ | $f_9$ | $f_{11}$ |
|------|-------|-------|-------|-------|-------|----------|
| 0.1  | 0.010 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000    |
| 0.2  | 0.038 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000    |
| 0.4  | 0.132 | 0.004 | 0.000 | 0.000 | 0.000 | 0.000    |
| 0.6  | 0.238 | 0.027 | 0.002 | 0.000 | 0.000 | 0.000    |
| 0.8  | 0.322 | 0.087 | 0.015 | 0.002 | 0.000 | 0.000    |
| 1.0  | 0.368 | 0.179 | 0.055 | 0.015 | 0.004 | 0.001    |
| 1.2  | 0.381 | 0.276 | 0.128 | 0.051 | 0.019 | 0.007    |
| 1.4  | 0.371 | 0.354 | 0.219 | 0.118 | 0.059 | 0.028    |
| 1.8  | 0.320 | 0.423 | 0.371 | 0.286 | 0.206 | 0.142    |
| 2.0  | 0.290 | 0.423 | 0.413 | 0.354 | 0.285 | 0.220    |
| 5.0  | 0.071 | 0.139 | 0.188 | 0.228 | 0.261 | 0.290    |
| 10.0 | 0.019 | 0.037 | 0.051 | 0.068 | 0.075 | 0.085    |
| 20.0 | 0.005 | 0.010 | 0.013 | 0.016 | 0.019 | 0.022    |

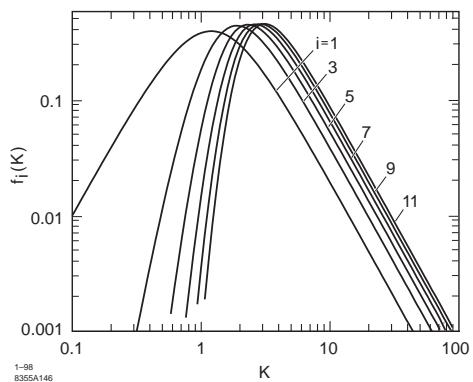


Figure 2: Functions  $f_i(K) = \frac{i^2 K^2 [JJ]}{(1 + \frac{1}{2}K^2)^2}$ .

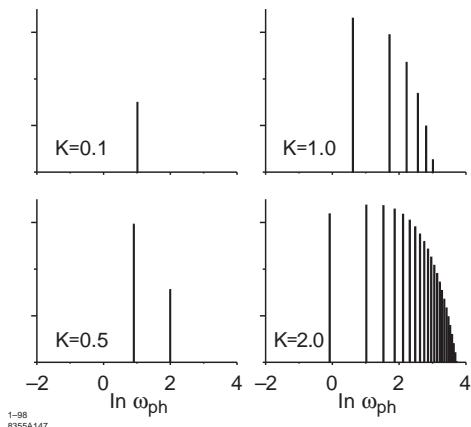


Figure 3: Line spectra of wiggler/undulator radiation for increasing strength parameter  $K$ .

Total spectral photon flux of the harmonic  $i$  into forward cone of angle  $\pm\sigma_{u'}$  is

$$\Phi_{ph}(\omega)|_i = \pi\alpha N_p \frac{\Delta\omega}{\omega} \frac{iK^2 [JJ]}{e 1 + \frac{1}{2}K^2} \quad (16)$$

$$= 1.431 \times 10^{17} IN_p \frac{\Delta\omega}{\omega} g_i(K)$$

where

$$g_i(K) = \frac{iK^2 [JJ]}{(1 + \frac{1}{2}K^2)} \quad (17)$$

Fig.3 shows the line spectrum as a function of  $K$ . For  $K < 1$  only few harmonics are obvious while for  $K \gg 1$  the line spectrum approaches the continuous synchrotron radiation spectrum of a wiggler or bending magnet.

The critical photon energy of wiggler radiation depends on the observation angle  $\psi$  in the deflecting plane because of a varying magnetic field,

$$\varepsilon_c(\psi) = \varepsilon_c(0) \sqrt{1 - \left(\frac{\gamma\psi}{K}\right)^2}. \quad (18)$$

## References

- [1] H. Wiedemann, Particle Accelerator Physics, 3rd Edition, Springer, Berlin Heidelberg 2007
- [2] K. Halbach, J. de Physique, C1, suppl.no 2, Tome 44 (1983)
- [3] H. Motz, JAP, vol. 22, no. 5 (1951) 527

### 3.1.5.1 Polarization of synchrotron radiation

H. Wiedemann, Stanford U./SSRL

Polarization of synchrotron radiation occurs in the direction of acceleration and radiation from bending magnet deflecting the beam in the horizontal plane is polarized to a high degree in the horizontal plane. This polarization is also known as the  $\sigma$ -polarization (orthogonal or in German "senkrecht" to magnetic field) in contrast to  $\pi$ -mode polarization which is parallel to the magnetic field. For intensities of both components see Sec.3.1.3.

Due to variations of source characteristics and observation, a variety of polarization conditions can be generated. A radiation field at the observer  $E(t) = (E_x(t), E_y(t), 0)$  has a generally complex spectral amplitude of  $A(\omega) = \int E(t)e^{i\omega t} dt = (A_x(\omega), A_y(\omega), 0)$  and an intensity  $I(\omega) = |A(\omega)|^2$ . With these definitions, the Stokes parameters [1] are defined by

$$\begin{aligned} S_0 &= I_x + I_y = A_x A_x^* + A_y A_y^* \\ S_1 &= I_x - I_y = A_x A_x^* - A_y A_y^* \\ S_2 &= I_{45} - I_{-45} = A_x A_y^* + A_y A_x^* \\ S_3 &= I_R - I_L = i(A_y A_x^* - A_x A_y^*) \end{aligned}$$

where  $I_{45}, I_{-45}$  are the intensities with polarization of  $\pm 45^\circ$ ;  $I_R, I_L$  are the intensities of right and left handed polarization; and  $A_{45}, A_{-45} = (A_x \pm A_y)/\sqrt{2}$ ;  $A_R, A_L = (A_x \pm iA_y)/\sqrt{2}$ , respectively. In ordinary synchrotron radiation  $A_y = 0$  and  $S_0 = S_1 = I_x$  while  $S_2 = S_3 = 0$ .

The electric field of synchrotron radiation from a positron is proportional to  $E \propto [\mathbf{n} \times (\mathbf{n} \times \dot{\beta})] = \mathbf{n}(\mathbf{n}\dot{\beta}) - \dot{\beta} = [n_x(\mathbf{n}\dot{\beta}) - \dot{\beta}_x, n_y(\mathbf{n}\dot{\beta}) - \dot{\beta}_y, n_z(\mathbf{n}\dot{\beta}) - \dot{\beta}_z]$ , or with  $\mathbf{n}\dot{\beta} = n_x\dot{\beta}_x + n_y\dot{\beta}_y + n_z\dot{\beta}_z$  and  $n_z \approx 1$  we get  $(E_x, E_y, 0) \approx (-\dot{\beta}_x, -\dot{\beta}_y + n_y n_x \dot{\beta}_x + n_y \dot{\beta}_z, 0)$ . Here,  $\mathbf{n}$  is the unit vector from observer to the source. On-axis,  $E_y = 0$  and we get only linearly polarized radiation. Off-axis ( $n_y \neq 0$ ) a vertical field component  $(E_y \approx n_y \dot{\beta}_z)$  appears for  $n_x = 0$ . Looking from above the midplane on an arc section of the particle path, we observe on the far away part of the arc a positive  $\dot{\beta}_z > 0$  and at the nearby part of the arc a negative acceleration  $\dot{\beta}_z < 0$  with an associated antisymmetric vertical electric field while the horizontal electric field is symmetric along this

### Sec.3.1: SYNCHROTRON RADIATION

arc section. The amplitude for  $n_y > 0$  is then  $\mathbf{A}(\omega) = \int (E_x \cos \omega t + i E_y \sin \omega t) dt$  or  $A_x$  is real and  $A_y$  is imaginary. For bending magnet radiation,  $S_2 = 0$  and  $S_3 \neq 0$  and depending on the sign of  $S_3$ , we have right ( $S_3 > 0$ ) or left handed ( $S_3 < 0$ ) circularly polarized radiation.

In wiggler and undulator magnets, we combine the radiation from a pair of adjacent poles and notice that the radiation is equal from both poles except for the sign of the acceleration. Therefore, the property of elliptically polarized radiation is cancelled. Ordinary flat wigglers and undulators exhibit only linear polarization in the horizontal plane.

#### Insertion devices for elliptically polarized radiation:

a) In an **asymmetric wiggler** field one can disturb the above mentioned cancellation by large and short positive fields followed by lower and longer negative fields. Such a magnet produces elliptically polarized radiation off-axis like that from bending magnets.

b) A **helical undulator** includes the fields  $(B_{0x} \sin k_p z, B_{0y} \cos k_p z)$  causing accelerations  $(\dot{\beta}_x, \dot{\beta}_y)$  and the fields on-axis are  $(E_x, E_y, 0) \approx (-\dot{\beta}_x, -\dot{\beta}_y, 0)$  generating circularly polarized radiation. Here  $\langle S_2 \rangle = 0$  and  $\langle S_3 \rangle \neq 0$ , indicating circularly or elliptically polarized radiation at a fundamental wavelength of  $\lambda = \frac{\lambda_p}{2\gamma^2} (1 + K^2)$  for  $B_{0x} = B_{0y}$  and  $\lambda = \frac{\lambda_p}{2\gamma^2} (1 + \frac{1}{2}K_x^2 + \frac{1}{2}K_y^2)$  for  $B_{0x} \neq B_{0y}$ .

c) **Cross field undulators:** Specific insertion devices of a great variety have been designed to produce a variation of polarization characteristics similar to the helical device.

d) **Variable polarization undulator:** Here the fields can be shifted in phase against each other  $[B_{0x} \sin(k_p z), B_{0y} \cos(k_p z + \Delta)]$  or  $E_x \propto \dot{\beta}_x \propto B_{0y} \cos(k_p z + \Delta)$  and  $E_y \propto \dot{\beta}_y \propto B_{0x} \sin(k_p z)$ . From that  $A_x \propto B_{0y} (\cos \Delta - i \sin \Delta)$  and  $A_y \propto B_{0x}$ . With this and for low values of  $K$  the Stokes parameters are for the fundamental wavelength  $S_1 \propto B_{0y}^2 - B_{0x}^2$ ,  $S_2 \propto 2B_{0x}B_{0y} \cos \Delta$ , and  $S_3 \propto 2B_{0x}B_{0y} \sin \Delta$ . Variation of  $\Delta$  by mechanical or electrical means allows the free choice of polarization.

#### References

- [1] G.G. Stokes, Trans.Cambridge Phil.Soc. 9 (1852) 399

#### 3.1.6 Transition and Diffraction Radiation

C. Thongbai, Chiang Mai U./ThEP

**Transition radiation** (TR), first predicted by Ginzberg and Frank [1], is emitted when a charged particle passes through an interface between two media with different dielectric constants. The angular spectral energy density for backward transition radiation emitted by an electron travelling from a medium with dielectric constant  $\varepsilon_1$  to  $\varepsilon_2$  [2]

$$\frac{dW}{d\Omega d\omega} = \frac{r_e m_e c^2 \beta^2 \varepsilon_1^{1/2} \sin^2 \theta_1 \cos^2 \theta_1}{\pi^2 c (1 - \beta^2 \varepsilon_1 \cos^2 \theta_1)} \times |F_1 F_2|^2 . \quad (1)$$

$$F_1 = \frac{(\varepsilon_1 - \varepsilon_2)}{(1 + \beta \sqrt{\varepsilon_2 - \varepsilon_1 \sin^2 \theta_1})} .$$

$$F_2 = \frac{(1 - \beta^2 \varepsilon_1 + \beta \sqrt{\varepsilon_2 - \varepsilon_1 \sin^2 \theta_1})}{(\varepsilon_2 \cos \theta_1 + \sqrt{\varepsilon_1 \varepsilon_2 - \varepsilon_1^2 \sin^2 \theta_1})} .$$

For forward radiation, exchange subscript 1 and 2 and replace  $\beta$  by  $-\beta$ . Applied to the transition from vacuum ( $\varepsilon_1 = 1$ ) to metal ( $\varepsilon_2 \rightarrow \infty$ ), the radiation energy becomes

$$\frac{dW}{d\Omega d\omega} = \frac{r_e m_e c^2 \beta^2 \sin^2 \theta_1}{\pi^2 c (1 - \beta^2 \cos^2 \theta_1)} . \quad (2)$$

The radiation intensity increases from zero at  $\theta_1 = 0$  to a broad peak at an angle  $\sin \theta_1 = 1/(\beta\gamma)$ .

For a  $\psi$ -incidence on a thin metal foil placed on the  $xy$ -plane, the forward radiation is emitted from the metal-to-vacuum interface with the radiation cone in the direction of the electron (at an angle  $\psi$  to the  $z$  axis). The backward radiation is emitted from vacuum-to-metal interface with the radiation cone in the direction of an angle  $\pi - \psi$  with respect to the  $z$  axis. Its spectral-angular distribution is given by the contribution of parallel and perpendicular polarization radiation with

$$\frac{dW^\parallel}{d\Omega d\omega} = \frac{r_e m_e c^2 \beta^2 \cos^2 \psi}{\pi^2 c [\sin \theta - \beta \cos \varphi \sin \psi]^2} \quad (3)$$

$$[(1 - \beta \sin \theta \cos \varphi \sin \psi)^2 - \beta^2 \cos^2 \theta \cos^2 \psi]^2 .$$

$$\frac{dW^\perp}{d\Omega d\omega} = \frac{r_e m_e c^2 \beta^2 \cos^2 \psi}{\pi^2 c [\beta \cos \theta \sin \varphi \sin \psi]^2} \quad (4)$$

$$[(1 - \beta \sin \theta \cos \varphi \sin \psi)^2 - \beta^2 \cos^2 \theta \cos^2 \psi]^2 .$$

The radiation angular distribution for oblique-incidence is azimuthally asymmetric. The asymmetry, however, vanishes for highly relativistic

electrons as the distribution becomes closer to that for normal incidence. Further details and discussions of other interface conditions can be found in [2, 3].

Although the radiation has its maximum intensity around an angle  $1/\gamma$ , most of the radiation is emitted at larger angles ( $> 96\%$  for  $\gamma > 10$ ) and collection of the radiation within an experimental acceptance angle is therefore important. Collectable radiation energy within an acceptance angle  $\theta_a$  is

$$\frac{dW}{d\omega}(\theta_a) = \frac{r_e m_e c^2}{2\pi c} \times \left[ \frac{2(1-\beta^2) \cos \theta_a}{(1-\beta^2 \cos^2 \theta_a)} + \frac{(1+\beta^2)}{\beta} \ln \left( \frac{(1-\beta \cos \theta_a)(1+\beta)}{(1+\beta \cos \theta_a)(1-\beta)} \right) - 2 \right].$$

While the total radiation emitted into half space is

$$\frac{dW}{d\omega} = \frac{r_e m_e c^2}{2\pi c} \left[ \frac{(1+\beta^2)}{\beta} \ln \left( \frac{(1+\beta)}{(1-\beta)} \right) - 2 \right], \quad (5)$$

which for relativistic electrons is proportional to  $\ln \gamma$ .

As for the spectral radiation distribution, the transition radiation spectrum from a single electron is uniform up to a very high frequency of around  $\gamma c \sqrt{4\pi Z N r_e}$ , where  $N$  is the atomic density and  $Z$  is the atomic number. For coherent transition radiation, this uniform spectrum folds with the form factor of the electron distribution in a bunch.

**Diffraction radiation** (DR) is emitted when a charged particle travels in the neighborhood of some inhomogeneity. Specifically, the radiation can be generated when an electron travels past a metallic structure such as a transition in beam line cross sections. The theory of DR was developed in the late 1950's, examining charged particles passing through circular apertures [4]. The radiation emitted from an electron moving through the middle of a circular aperture of radius  $a$  in an ideal conducting screen can be expressed as

$$\frac{dW}{d\omega d\Omega} = \frac{dW_{TR}}{d\omega d\Omega} \alpha^2 \left[ J_0 \left( \frac{\omega a}{c} \sin \theta \right) \left( \frac{\omega a}{c\beta\gamma} \right) K_1 \left( \frac{\omega a}{c\beta\gamma} \right) \right]^2. \quad (6)$$

Here,  $dW_{TR}/d\omega d\Omega$  is the transition radiation defined in Eq.(2),  $\theta$  is the observation angle with respect to the beam axis,  $J_0$  is the Bessel function of order zero and  $K_1$ ,  $I_1$  are modified Bessel functions of the first order. For backward DR,  $\theta$  is the

angle between the radiation direction and  $-\vec{v}$ . For a single electron,  $\alpha = 1$  and the radiated intensity approaches that of transition radiation when the aperture radius becomes very small ( $a \rightarrow 0$ ).

For a uniform beam of radius  $b$ ,  $\alpha = (2c\beta\gamma/\omega b)I_1(\omega b/c\beta\gamma)$ .

For an electron passing the aperture at a distance  $r$  from the center [1], add the term  $dW_{TR}/d\omega d\Omega [(r/a) J_1(\omega a \sin \theta/c)]^2$  to Eq.(6).

For a discussion of diffraction radiation from a slit, see [2, 5] and references therein.

## References

- [1] V. Ginzberg, I. Frank, J.Phys. 9 (1945) 353
- [2] M.L. Ter-Mikaelian, High-energy Electromagnetic Process in Condensed Media, Wiley-Interscience, New York, 1972
- [3] V.L. Ginzburg, V.N. Tsytovich, Transition Radiation and Transition Scattering, Adam Hilger, Bristol, 1990
- [4] Yu.N. Dnestrovskii, D.P. Kostomorov, Sov. Phys. Dokl. 4 (1959) 132; Sov. Phys. Dokl. 4 (1959) 158
- [5] M. Castellano, NIM A394 (1977) 275

### 3.1.7 Coherent Synchrotron Radiation

*S. Krinsky, BNL*

(See also Secs.2.4.15, 3.1.12) Consider a bunch containing  $N$  highly relativistic electrons passing through a magnet with bending radius  $\rho$ . The power radiated per unit wavelength  $P(\lambda)$  is [1]

$$P(\lambda) = p(\lambda) [N + N(N-1)g(\lambda)], \quad (1)$$

where  $p(\lambda)$  is the power radiated per unit wavelength for a single electron and  $g(\lambda)$  is the absolute square of the Fourier transform of the bunch distribution. For a Gaussian bunch of rms length  $\sigma$ ,  $g(\lambda) = \exp(-4\pi^2\sigma^2/\lambda^2)$ . The first term in Eq.(1) with linear dependence on  $N$  is the incoherent power and the second term with quadratic dependence is the coherent power. Coherent synchrotron radiation (CSR) has been observed using linacs [2] and storage rings [3, 4]. For a Gaussian bunch in free space, the total coherent power radiated is

$$P_{coh}^{FS}[W] = 2.42 \times 10^{-20} \frac{N^2}{\rho^{2/3} [m] \sigma^{4/3} [m]}. \quad (2)$$

In the case of electrons moving in the midplane between parallel conducting plates with full separation  $h$ , the single electron power spectrum is cut off for wavelength greater than the cutoff

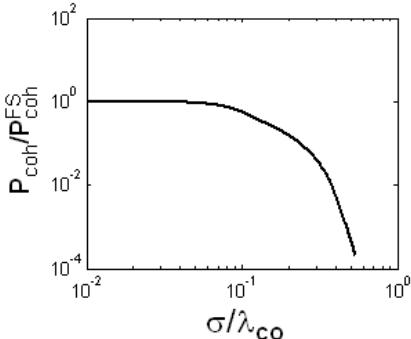


Figure 1: For a Gaussian bunch, the ratio of the coherent power radiated between parallel plates to that in free space is plotted versus the ratio of the rms bunch length to the cutoff wavelength.

wavelength  $\lambda_{co} = 2h(h/\rho)^{1/2}$  [5] according to  $p(\lambda) \propto \exp(-\lambda^2/2\lambda_{co}^2)$ . The total radiated power is reduced relative to that radiated in free space by the factor  $P_{coh}/P_{coh}^{FS} = F(\sigma/\lambda_{co})$  [6] shown in Fig.1. We see that if  $\sigma/\lambda_{co} < 0.1$ , then the total coherent power is very closely equal to that in free space, but for larger bunch length, the total coherent power drops rapidly.

Coherent emission results from the individual electrons moving in the electric fields radiated by other electrons. In accelerator physics this type of phenomena is generally treated using the formalism of wakefields or impedance. In addition to enhancing the radiated power, the CSR impedance can result in an alteration of the electron bunch distribution. In a storage ring, the CSR impedance can give rise to the microwave instability (Sects.2.4.9, 2.4.15). This instability can result in the generation of a high-frequency density modulation of the bunch and in an increase of the CSR power at the modulation frequencies. One problem in utilizing this mechanism for a radiation source is that it is not a reproducible steady state phenomenon. The stability condition for the microwave instability driven by the CSR impedance is [7]

$$\frac{I_p}{I_A} < \frac{\gamma \alpha_p \delta_0^2 C}{2(\pi \rho \lambda_{co}^2)^{1/3}}, \quad (3)$$

where  $\gamma$  is the relativistic factor,  $\alpha_p$  is the momentum compaction,  $\delta_0$  is the low current energy spread,  $C$  is the storage ring circumference,  $I_p$  is the peak current of the bunch and  $I_A = ec/r_e = 17,000$  A.

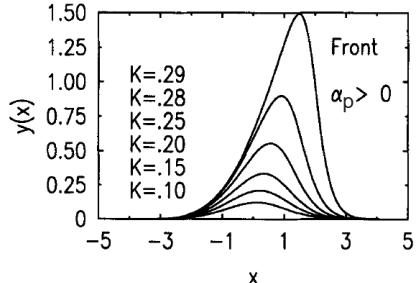


Figure 2: Equilibrium beam distribution for CSR wake and positive momentum compaction.

At low current in a storage ring, the bunch distribution is Gaussian. At higher currents, the CSR impedance can result in potential well distortion, which modifies the bunch distribution. Below the microwave threshold, this is a reproducible steady-state phenomenon that for positive momentum compaction sharpens the fall of the bunch density at the front of the bunch and extends the CSR power spectrum to higher frequencies. The modification of the bunch distribution is described by the Haissinski equation. We define the dimensionless variables,  $x = s/\sigma_0$  and  $y(x) = (Z_0/\sigma_0 V'_{rf})(\rho/3\sigma_0)^{1/3} I(s)$ , where  $\sigma_0$  is the rms length of the low-current Gaussian bunch,  $V'_{rf} = dV_{rf}/ds$  is the spatial derivative of the rf voltage, and  $I(s)$  is the current distribution. The step-function CSR wake potential is given by  $S(s) = -Z_0(\rho/3s)^{1/3}$ . The Haissinski equation is [8, 9]

$$y(x) = K \exp \left[ -\frac{x^2}{2} + \text{sgn}(\alpha_p) \int_0^\infty dz y(z) z^{-1/3} \right], \quad (4)$$

where the constant  $K$  is determined by the normalization condition  $\int ds I(s) = Nec$  and hence depends on the number of electrons in the bunch. For positive momentum compaction, the distortion of the bunch distribution is illustrated in Fig.2.

## References

- [1] J.S. Nodwick, D.S. Saxon, PR 96 (1954) 180
- [2] T. Nakazato et al, PRL 63 (1989) 1245
- [3] G.L. Carr et al, NIM A463 (2001) 387

- [4] M. Abo-Bakr et al, PRL 88, 254801 (2002); 90, 094801 (2003)
- [5] R.L. Warnock, SLAC-PUB-5375 (1990)
- [6] J.B. Murphy et al, PA 57 (1997) 9
- [7] G. Stupakov, S. Heifets, PRST-AB 5, 054402 (2002)
- [8] K. Bane et al, AIP Conf. Proc. 367 (1995) p.191
- [9] F. Sannibale et al, PRL 93, 094801 (2004)

### 3.1.8 Free-Electron Lasers

*Z. Huang, SLAC  
P. Schmüser, U. Hamburg*

**Overview** The free-electron laser (FEL) is an accelerator-based coherent light source of extreme brilliance. In contrast to conventional lasers, the electrons in an FEL are not bound to matter, but travel as free particles at relativistic speed ( $\beta = v/c \approx 1$ ,  $\gamma = 1/\sqrt{1-\beta^2} \gg 1$ ) through an undulator magnet. The light wavelength is inversely proportional to the square of the electron energy and can be tuned at will. FELs may operate as oscillators, starting from spontaneous radiation, or as amplifiers, boosting the light from a seed laser. FELs in the THz, infrared and visible regime are equipped with an optical cavity and work in the low-gain regime. At ultraviolet and x-ray wavelengths, optical cavities are presently not available and a large FEL gain must be achieved in a single passage through a very long undulator magnet. Of particular importance is the Self-Amplified Spontaneous Emission (SASE) process. SASE FELs are capable of producing coherent ultraviolet and x-ray pulses of femtosecond duration with a peak brilliance  $10^6 - 10^{10}$  times higher than in third-generation synchrotron light sources. For a systematic presentation of FEL theory see Refs.[1, 2, 3, 4] and the literature quoted therein.

**Undulator Radiation** The undulator axis is chosen along the beam direction  $z$ , the magnetic field in  $y$  direction (vertical). The magnet period  $\lambda_u$  is a few cm. The on-axis field in a planar undulator is nearly harmonic,  $B_y = -B_0 \sin(k_u z)$ , with  $k_u = 2\pi/\lambda_u$ . The electrons travel on a sine-like trajectory  $x(z) = K/(\beta\gamma k_u) \sin(k_u z)$  with the *undulator parameter*

$$K = eB_0\lambda_u/(2\pi m_e c), \quad K \approx 1. \quad (1)$$

The transverse and longitudinal velocities are

$$v_x(z) = K c/\gamma \cos(k_u z) \quad (2)$$

$$v_z \approx c \left(1 - (1 + K^2/2)/(2\gamma^2)\right) \quad (3)$$

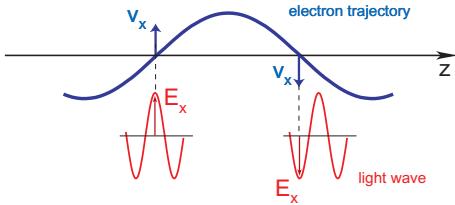


Figure 1: Condition for sustained energy transfer from electron to light wave.

The fundamental wavelength of undulator radiation in forward direction is

$$\lambda_\ell = \frac{\lambda_u}{2\gamma^2} \left(1 + \frac{K^2}{2}\right). \quad (4)$$

**Low-Gain FEL Theory** A low-gain FEL consists of a storage ring or linac, a short undulator, and an optical cavity. Upon each passage through the undulator the light intensity grows by only a few per cent, but after very many round trips a large average FEL beam power can be achieved, e.g. more than 10 kW in the infrared FEL at JLAB [5].

The light wave co-propagating with the electron beam is taken as a plane wave  $E_x(z, t) = E_0 \cos(k_\ell z - \omega_\ell t + \psi_0)$  with wavelength  $\lambda_\ell$  and wave number  $k_\ell = 2\pi/\lambda_\ell = \omega_\ell/c$ . The transverse electron velocity  $v_x$  and the field  $E_x$  must point in the same direction to enable energy transfer. Sustained energy transfer from electron to light wave requires that the light wave slips forward with respect to the electron by  $\lambda_\ell/2$  per half period of the electron trajectory, see Fig.1. This is fulfilled if  $\lambda_\ell$  and  $\gamma$  are related by Eq.(4). Slippages by  $3\lambda_\ell/2, 5\lambda_\ell/2 \dots$  are also permitted, leading to odd higher harmonics ( $\lambda_\ell/3, \lambda_\ell/5 \dots$ ) of the FEL radiation. However slippages of  $2\lambda_\ell/2, 4\lambda_\ell/2 \dots$  yield zero net energy transfer, hence even harmonics are absent.

The time derivative of the electron energy is

$$m_e c^2 d\gamma/dt = -ev_x E_x \\ = -eCKE_0/(2\gamma) [\cos \psi + \cos \chi] \quad (5)$$

with  $\psi = (k_\ell + k_u)z - \omega_\ell t + \psi_0$ .  $\psi$  is called *ponderomotive phase*, and  $\chi = \psi - 2k_u z$ . If Eq.(4) holds,  $\cos \psi$  remains constant along the undulator, while  $\cos \chi$  is rapidly oscillating and cancels out. Choosing  $\psi = 0, \pm 2\pi, \pm 4\pi \dots$  yields optimum energy transfer along the entire electron trajectory.

**FEL pendulum equations.** Let the lasing process be started by light of wavelength  $\lambda_\ell$ . The

### Sec.3.1: SYNCHROTRON RADIATION

resonance electron energy  $\gamma_r m_e c^2$  and the relative energy deviation  $\eta$  are defined by

$$\gamma_r = \sqrt{\frac{\lambda_u (1 + K^2/2)}{2\lambda_\ell}}, \quad \eta = \frac{\gamma - \gamma_r}{\gamma_r} \quad (6)$$

The FEL pendulum equations read

$$\psi' = d\psi/dz = 2k_u \eta, \quad (7)$$

$$\eta' = d\eta/dz = -eE_0 \hat{K}/(2m_e c^2 \gamma_r^2) \cos \psi. \quad (8)$$

In a planar undulator,  $v_z$  contains an oscillatory term which leads to odd higher harmonics and influences the energy transfer. This is accounted for by modifying the undulator parameter in Eq.(8)

$$\hat{K} = K [J_0(\xi) - J_1(\xi)], \quad \xi = K^2/(4 + 2K^2). \quad (9)$$

The FEL gain function, defined as the relative increase in light intensity during one passage of the undulator,  $G = \Delta I/I$ , is proportional to the negative derivative of the line-shape curve of undulator radiation (Madey theorem) [6]. Electrons with  $\eta > 0$  enhance the light intensity, those with  $\eta < 0$  reduce it, see Fig.4 below.

**1D High-Gain FEL Theory** High-gain FEL radiation has a much higher intensity than undulator radiation because many electrons radiate coherently. *Microbunching* is essential: electrons losing energy to the light wave travel on a sinusoidal trajectory of larger amplitude than electrons gaining energy. The resulting modulation of  $v_z$  leads to the formation of microbunches. The particles in a microbunch radiate like a single particle of high charge. The growing radiation field enhances the microbunching further and leads to an exponential growth and eventual saturation of the radiation power. Results from LCLS (*Linac Coherent Light Source*) are shown in Fig.2 [7].

Wave equation. In the one-dimensional (1D) FEL theory the dependencies of bunch current and electromagnetic fields on the transverse coordinates  $x, y$  are neglected. The radiation field obeys the 1D inhomogeneous wave equation

$$\left[ \frac{\partial^2}{\partial z^2} - \frac{1}{c^2} \frac{\partial^2}{\partial t^2} \right] \tilde{E}_x(z, t) = \mu_0 \frac{\partial \tilde{j}_x}{\partial t}, \quad (10)$$

where  $\tilde{j}_x$  is transverse current density resulting from the sinusoidal motion. We make the ansatz

$$\tilde{E}_x(z, t) = \tilde{E}_x(z) \exp[i(k_\ell z - \omega_\ell t)] \quad (11)$$

with a complex amplitude function  $\tilde{E}_x(z)$ . The microbunching effect is anticipated by assuming a small periodic modulation of the current density

$$\tilde{j}_x(\psi, z) \approx \frac{K}{\gamma} \cos(k_u z) [j_0 + \tilde{j}_1(z) \exp(i\psi)] \quad (12)$$

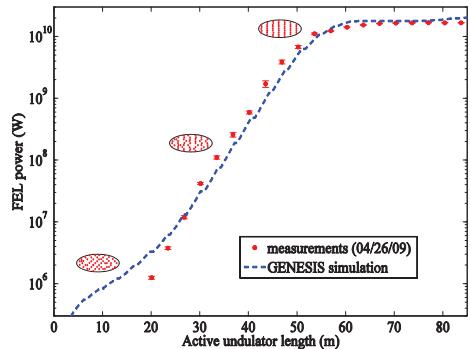


Figure 2: Exponential growth and saturation of the FEL power in LCLS at  $\lambda = 1.5 \text{ \AA}$  as a function of active undulator length [7]. The progressing microbunching is indicated schematically.

with a slowly growing complex amplitude  $\tilde{j}_1(z)$ . Combining Eqs.(10), (11) and (12) yields

$$\tilde{E}'_x(z) = d\tilde{E}_x/dz = -\mu_0 c \hat{K}/(4\gamma_r) \tilde{j}_1(z). \quad (13)$$

Coupled first-order equations. To study the phase space motion of many electrons, the bunch is subdivided into slices. In the slice  $0 \leq \psi < 2\pi$  we choose  $N$  electrons which phases  $\psi_n$ . In the *periodic model* the distribution is continued in the other slices. Then from Fourier series expansion

$$\tilde{j}_1 = 2j_0 \sum_{n=1}^N \exp(-i\psi_n)/N. \quad (14)$$

$\sum \exp(-i\psi_n)/N$  is called *bunching factor*. Combining Eqs.(7), (8) and (13) one obtains a set of coupled first-order equations

$$\psi'_n = 2k_u \eta_n, \quad n = 1 \dots N \quad (15)$$

$$\eta'_n = -e \hat{K}/(2m_e c^2 \gamma_r^2) \operatorname{Re} \left\{ \tilde{E}_x \exp(i\psi_n) \right\}$$

$$\tilde{E}'_x = -\mu_0 c \hat{K}/(4\gamma_r) \cdot \tilde{j}_1,$$

which, together with Eq.(14), describe the evolution of the phases  $\psi_n$  and energy deviations  $\eta_n$  of the  $N$  electrons, as well as the growth of  $\tilde{E}_x(z)$  and  $\tilde{j}_1(z)$ . Longitudinal space charge forces are of minor importance in short-wavelength FELs and are neglected here and in Eq.(16) below.

Third-order equation. The main physics of the high-gain FEL is contained in the first-order Eqs.(15) but these can only be solved numerically. If the modulation current  $\tilde{j}_1$  remains small a linear third-order differential equation for the elec-

tric field can be derived (see e.g. [4]):

$$\tilde{E}_x''' + 4ik_u\eta\tilde{E}_x'' - 4k_u^2\eta^2\tilde{E}_x' - \frac{i}{(\sqrt{3}L_{g0})^3}\tilde{E}_x = 0 \quad (16)$$

with the 1D power gain length

$$L_{g0} = \frac{1}{\sqrt{3}} \left[ \frac{4\gamma_r^3 m_e}{\mu_0 \hat{K}^2 e^2 k_u n_e} \right]^{1/3}. \quad (17)$$

Here  $n_e$  is the electron density.

Exponential gain and saturation. The solution of Eq.(16) is of the form  $\tilde{E}_x(z) = \sum_{j=1}^3 A_j \exp(\alpha_j z)$ . For the special case  $\eta = 0$  (electrons are on resonance) one finds  $\alpha_{1,2} = (\pm 1 + i/\sqrt{3})/(2L_{g0})$ ,  $\alpha_3 = -i/(\sqrt{3}L_{g0})$ . In case of laser seeding with an initial field  $E_0$ , all amplitudes are equal,  $A_j = E_0/3$ . The light power stays almost constant in the “*lethargy regime*”,  $0 \leq z \lesssim 2L_{g0}$ , but then it grows exponentially (see Fig.3)

$$P(z) \propto \exp(2Re[\alpha_1]z) \equiv \exp(z/L_{g0}). \quad (18)$$

The Eqs.(15) yield the same result in the lethargy and exponential gain regimes but describe FEL saturation in addition. The saturation power is

$$P_{\text{sat}} \approx \rho P_b, \quad (19)$$

where  $P_b$  the electron beam power, and  $\rho$  is the dimensionless FEL (Pierce) parameter [9]

$$\rho = \frac{\lambda_u}{4\pi\sqrt{3}L_{g0}} = \left[ \frac{\pi I_0}{8 I_A} \frac{\hat{K}^2}{\gamma_r^3 A_b k_u^2} \right]^{1/3}. \quad (20)$$

( $I_0$  peak current,  $I_A \approx 17$  kA Alfvén current,  $A_b$  beam cross section). For short-wavelength FELs,  $\rho \approx 10^{-3}$ .

FEL gain-function and bandwidth. For a short undulator (length  $\leq L_{g0}$ ), the high-gain FEL theory agrees perfectly with the low-gain theory, but in long undulators strong differences are seen: the gain is much larger and the gain-function approaches a Gaussian (Fig.4). The high-gain FEL acts as a narrow-band amplifier with an rms bandwidth

$$\sigma_\omega/\omega = \sqrt{0.83\rho\lambda_u/z}. \quad (21)$$

SASE [8, 9] permits the startup of lasing without seed radiation. Intuitively, spontaneous undulator radiation produced in the first section of a long undulator serves as seed radiation in the remaining section. Because of the random electron distribution, the current contains a noise term

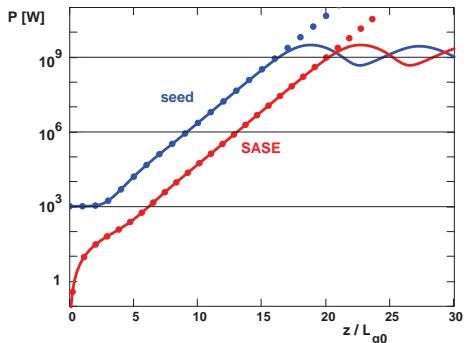


Figure 3: FEL power as a function of  $z/L_{g0}$  in a seeded FEL (blue) and a SASE FEL (red). Solid curves: numerical integration of coupled first-order Eqs.(15). Dots: analytic solution of third-order Eq.(16).

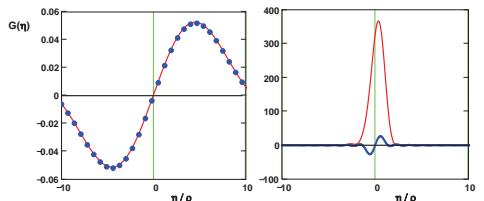


Figure 4: FEL gain function  $G(\eta)$  plotted vs.  $\eta/\rho$  at two positions in a long undulator:  $z = 1L_{g0}, 8L_{g0}$ . Red curves: high-gain theory. Blue dots/curve: low-gain theory (Madey theorem).

which has a spectral component within the FEL bandwidth. The effective shot-noise power and modulated current density are [10, 4]

$$P_n = \rho \gamma m_e c^2 \sigma_\omega / (2\pi), \quad \tilde{j}_1(0) \approx \sqrt{e I_0 \sigma_\omega} / A_b \quad (22)$$

The computed power rise for typical parameters of the soft x-ray FEL FLASH (see e.g. [4]) is shown in Fig.3. Saturation is achieved at an undulator length  $L_u \approx 20L_{g0}$ . The SASE bandwidth at saturation is  $\sigma_\omega^s/\omega \approx \rho$ .

SASE radiation exhibits shot-to-shot fluctuations in wavelength. The coherence length at saturation is

$$L_{\text{coh}} \approx \sqrt{\pi}c/\sigma_\omega^s = \lambda_\ell / (2\sqrt{\pi}\rho). \quad (23)$$

For a bunch length  $L_b > L_{\text{coh}}$ , the average number of spikes in the wavelength spectra is  $M = L_b/L_{\text{coh}}$ . In the exponential gain regime the normalized radiation pulse energy  $u = U_{\text{rad}}/\langle U_{\text{rad}} \rangle$

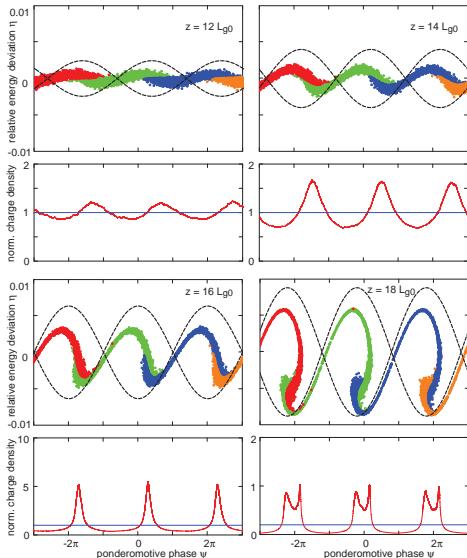


Figure 5: Evolution of the microbunch structure at  $z = 12L_{g0}, 14L_{g0}, 16L_{g0}, 18L_{g0}$ . Upper subplots: distribution of particles in  $(\psi, \eta)$  phase space. Three FEL buckets are indicated by dashed curves. Lower subplots: normalized charge density as function of  $\psi$ .

fluctuates according to the *gamma distribution*

$$p_M(u) = \frac{M^M u^{M-1}}{\Gamma(M)} e^{-Mu}, \quad \sigma_u^2 = 1/M. \quad (24)$$

Phase space and simulation of microbunching. The FEL dynamics resembles the synchrotron oscillations of a stored proton beam. In the  $(\psi, \eta)$  phase space the particles rotate clockwise, hence particles in the right half of an FEL bucket transfer energy to the light wave, while those in the left half withdraw energy. Eqs.(15) are well suited for modelling the microbunching, see Fig.5. For  $z \geq 12L_{g0}$  pronounced microbunches evolve in the right halves of the FEL buckets and increase the light intensity, while beyond  $18L_{g0}$  they move into the left halves and reduce it. The FEL power oscillations in Fig.3 are caused by this rotation in phase space.

Higher harmonics. Close to saturation, the periodic sequence of narrow microbunches (see Fig.5) corresponds to a modulation current with rich harmonics contents. In a planar undulator, odd higher harmonics will be amplified. The third

(fifth) harmonic can reach 1% (0.1%) of the fundamental power.

**3D Effects** The realistic description of high-gain FELs has to be based on a three-dimensional (3D) theory, taking into account electron beam emittance and energy spread, and optical diffraction. For idealized cases, e.g. round beam with uniform longitudinal charge density, an FEL eigenmode equation including all these effects can be developed. Realistic cases require sophisticated simulation codes such as FAST [12], GENESIS [13] or GINGER [14]. These are indispensable for the design of short-wavelength FELs.

3D gain length. The 3D gain length  $L_g$  is typically 30 – 50% longer than the 1D gain length  $L_{g0}$ . In [15]  $L_g$  is expressed in terms of three dimensionless parameters:  $X_\gamma = L_{g0} 4\pi\sigma_\eta/\lambda_u$  (energy spread par.),  $X_d = L_{g0}\lambda_\ell/(4\pi\sigma_r^2)$  (diffraction par.,  $\sigma_r$  rms beam radius) and  $X_\varepsilon = L_{g0} 4\pi\varepsilon/(\beta_{av}\lambda_\ell)$  (angular spread par.,  $\varepsilon$  emittance,  $\beta_{av}$  average beta function).

$$L_g = L_{g0}(1 + \Lambda) \quad (25)$$

$$\begin{aligned} \Lambda = & a_1 X_d^{a_2} + a_3 X_\varepsilon^{a_4} + a_5 X_\gamma^{a_6} + a_7 X_\varepsilon^{a_8} X_\gamma^{a_9} \\ & + a_{10} X_d^{a_{11}} X_\gamma^{a_{12}} + a_{13} X_d^{a_{14}} X_\varepsilon^{a_{15}} \\ & + a_{16} X_d^{a_{17}} X_\varepsilon^{a_{18}} X_\gamma^{a_{19}} \end{aligned}$$

$$a_1 = 0.45, \quad a_2 = 0.57, \quad a_3 = 0.55, \quad a_4 = 1.6,$$

$$a_5 = 3.0, \quad a_6 = 2.0, \quad a_7 = 0.35, \quad a_8 = 2.9,$$

$$a_9 = 2.4, \quad a_{10} = 51, \quad a_{11} = 0.95, \quad a_{12} = 3.0,$$

$$a_{13} = 5.4, \quad a_{14} = 0.7, \quad a_{15} = 1.9, \quad a_{16} = 1140,$$

$$a_{17} = 2.2, \quad a_{18} = 2.9, \quad a_{19} = 3.2.$$

Gain guiding counteracts the diffractive widening of the FEL beam since most of the light is generated in the central core of the electron beam [16]. Gain guiding permits the FEL beam to follow slow, “adiabatic” motions of the electron beam and is thus crucial for the long undulator in an x-ray FEL.

Transverse coherence. The fundamental Gaussian mode  $\text{TEM}_{00}$  has its highest intensity on the beam axis while higher modes extend to larger radial distances. The  $\text{TEM}_{00}$  mode grows fastest along the undulator, owing to its superior overlap with the electron beam. Near saturation it dominates and the FEL radiation possesses a high degree of transverse coherence, as verified by double-slit diffraction experiments [11].

**Seeding** Various seeding methods have been proposed to improve the coherence length of SASE radiation. Direct seeding requires a coherent signal well above the shot-noise level. In

the VUV such signals may be obtained by high harmonic generation (HHG) in a gas. At shorter wavelengths, *self-seeding* [17] may be applied: a SASE signal, produced in a short undulator, is passed through a monochromator and serves as seed radiation in the main undulator. In a high-gain harmonic generation (HGHG) FEL [18], the electron beam is energy-modulated in an undulator by interaction with a powerful laser. A magnetic chicane converts the energy modulation to a density modulation. A second undulator causes the density-modulated beam to emit coherent radiation at a higher harmonic frequency. In an echo-enabled harmonic generation (EEHG) FEL [19], a second modulator followed by a second chicane are inserted before the radiator. The electron beam interacts twice with two laser pulses in the two modulators. The longitudinal phase space becomes highly nonlinear, leading to density modulations at a very high harmonic number for a modest energy modulation.

**Beam requirements** Very bright electron beams are required to drive ultraviolet and x-ray FELs. Higher peak current and smaller cross section reduce the gain length (see Eq.(20)). High peak currents require longitudinal bunch compression (see Sec.4.5), but the energy spread is increased by this process (which affects the gain length through  $X_\gamma$  in Eq.(25)). Very low-emittance beams can be generated with photocathode rf guns or specially designed thermionic guns. The beam cross section in the undulator can be reduced by stronger focusing (i.e., smaller  $\beta_{av}$ ), but the increased angular spread will eventually degrade the FEL gain (through  $X_\varepsilon$  in Eq.(25)). The FEL design optimization is therefore multi-dimensional and beyond our scope here, but the typical requirements on electron beams are

$$I_0 \geq 1\text{ kA}, \quad \sigma_\eta \leq \rho/2, \quad \varepsilon \sim \lambda_\ell/(4\pi). \quad (26)$$

These requirements apply to the “slice” beam qualities defined on the scale of the coherence length (see Eq.(23)). For harmonic generation FELs, the slice energy spread should be much smaller than the  $\rho$ -parameter of the final amplifier because the additional energy modulation imposed on the beam becomes the effective energy spread there. Beam current, slice emittance and energy spread should be “flat” along the bunch in order not to increase the final radiation bandwidth. Such high-quality beams can be produced with linacs but not with storage rings.

## References

- [1] J.B. Murphy, C. Pellegrini, Introduction to the physics of free electron lasers, Laser Handbook Vol. 6 (1990) 115
- [2] E.L. Saldin, E.A. Schneidmiller, M.V. Yurkov, The Physics of Free Electron Lasers, Springer 2000
- [3] Z. Huang, K.-J. Kim, PRST-AB 30, 034801 (2007)
- [4] P. Schmüser, M. Dohlus, J. Rossbach, Ultraviolet and Soft X-Ray Free-Electron Lasers, Springer 2008
- [5] FEL at JLAB: see [www.jlab.org/FEL](http://www.jlab.org/FEL)
- [6] J.M.J. Madey, Nuovo Cim. 50B (1979) 64
- [7] P. Emma et al, PAC 09
- [8] A.M. Kondratenko, E.L. Saldin, PA 19 (1980) 207
- [9] R. Bonifacio, C. Pellegrini, M. Narducci, Opt. Comm. 50 (1984) 373
- [10] K.-J. Kim, PRL 57 (1986) 1871
- [11] R. Ischebeck et al, NIM A507 (2000) 175
- [12] E.L. Saldin et al, NIM A429 (1999) 233
- [13] S. Reiche, NIM A429 (1999) 243
- [14] W. Farlay, Report LBNL-49625, LBL (2002)
- [15] M. Xie, NIM A445 (2000) 59
- [16] G. Moore, NIM A239 (1985) 19
- [17] J. Feldhaus et al, Opt. Com. 140 (1997) 341
- [18] L.H. Yu, PR A44 (1991) 5178
- [19] G. Stupakov, PRL 102, 074801 (2009)

### 3.1.9 Ultrashort X-ray Pulse Generation

A. Zholents, ANL

Apart from squeezing the electron bunch [1, 2], the other way to obtain ultrashort x-ray pulses is to select the radiation produced by a small section of the electron bunch. In the case of spontaneous emission, this is accomplished by physically pointing this radiation into a different direction than the radiation of other electrons. In the case of coherent emission, this is accomplished by providing a condition when this radiation dominates the radiation of all other electrons by a large factor.

**Spontaneous emission** “Slicing” of the electron bunches has been implemented in storage rings [3] where electron beam interaction with a laser in a wiggler magnet has been used to produce energy modulation of electrons with modulation amplitude  $\Delta E$  several times larger than the electron beam energy spread  $\sigma$ :

$$\begin{aligned} \Delta E(q, \nu, \hat{\sigma}_\tau) = & 2\sqrt{A_L \alpha \hbar \omega_L \xi} (J_0(\xi/2) \\ & - J_1(\xi/2)) f(q, \nu, \hat{\sigma}_\tau) \end{aligned} \quad (1)$$

### Sec.3.1: SYNCHROTRON RADIATION

where

$$f(q, \nu, \hat{\sigma}_\tau) = \frac{1}{\sqrt{\frac{2q}{(2\pi)^{1/2}\hat{\sigma}_\tau}}} \int_{-0.5}^{0.5} \frac{\cos(2\pi\nu\hat{z} - \tan^{-1}(q\hat{z}))}{\sqrt{1+(q\hat{z})^2}} e^{-\hat{z}^2/4\hat{\sigma}_\tau^2} d\hat{z},$$

and  $\alpha$  is the fine structure constant;  $A_L$  the laser pulse energy;  $\omega_L = 2\pi c/\lambda_L$  with  $\lambda_L$  the laser wavelength;  $q = L_w/z_R$  with  $L_w$  the wiggler length and  $z_R$  the Rayleigh length;  $\sigma_\tau$  the rms width of the laser intensity distribution;  $\hat{z} = z/L_w$  with  $z$  the longitudinal coordinate centered at the wiggler center;  $\xi = K^2/(2+K^2)$ , where  $K = eB_0\lambda_W/(2\pi m_e c)$  with  $B_0$  the wiggler peak magnetic field and  $\lambda_W$  the wiggler period;  $\nu = N2(\delta\gamma/\gamma_r)$  is the detuning from the resonance energy with relativistic factor  $\gamma_r^2 = \lambda_W/2\lambda_L(1+K^2/2)$ ;  $\hat{\sigma}_\tau = \sigma_\tau c/N\lambda_L$ ; and  $N$  is the number of wiggler periods. Then lattice dispersion is used to separate the emission from electrons with large  $\Delta E$ . The actual source of the synchrotron radiation (SR) can be bending magnet, wiggler, or undulator. The duration of the x-ray pulse is defined by the laser pulse and by stretching of the electron slice accumulated on a path from the wiggler to the radiation source. The timing for the x-ray pulse is defined by the arrival of the laser pulse and, thus, is synchronized with the laser pulse.

Another technique uses a deflecting cavity in a storage ring to produce a time-dependent kick to the electron bunch that changes the sign in the middle of the electron bunch and reaches a large magnitude at the bunch edges which exceeds by many times the angular divergence of the x-ray emission in the undulator  $\sigma_{r'}$  and the rms beam angular spread  $\sigma_{y'}$  [4]. This kick is eventually canceled by a second deflecting cavity, but the x-ray beamline located between these two cavities can use the time-dependent variation imprinted on the x-ray pulse from the SR source either by selecting a short slice of the radiation with a slit or by compressing the entire radiation pulse using an asymmetrically cut crystal:

$$\sigma_{\tau-xray} \approx \frac{E_b}{eU} \frac{1}{2\pi f_{rf}} \sqrt{\frac{\varepsilon_y}{\beta_{rf}} \left( 1 + \frac{\sigma_{r'}^2}{\sigma_{y'}^2} \right)}. \quad (2)$$

Here  $U$  is the “deflecting voltage”,  $f_{rf}$  is the deflecting cavity frequency,  $E_b$  is the electron beam energy,  $\varepsilon_y$  is the vertical electron beam emittance, and  $\beta_{rf}$  is the vertical beta function at the deflecting cavity.

**Coherent emission** In the case of coherent emission from free-electron lasers (FELs), one

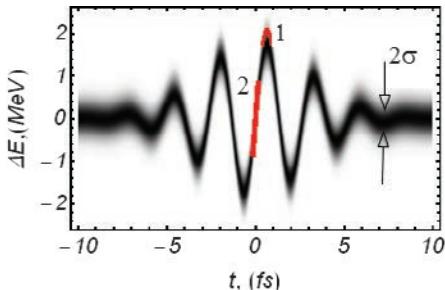


Figure 1: Density plot showing a fragment of the longitudinal phase space of the electron bunch after interaction with a few-cycle laser pulse in one period wiggler magnet. The width of the curve is defined by the electron energy spread. Two regions of interest (see text) are highlighted by red color and labelled (1) and (2).

can prepare the electron bunch entering the FEL undulator in such way that an ultrashort slice of the electron bunch will lase, and all the rest of the electrons will not lase or lase only with a much slower pace, thus producing a much weaker signal. One possibility is to use a slotted foil and spoil the emittance of the entire electron bunch except for a small fraction that goes through a slot in the foil [5]. Another possibility is to use a few-cycle laser pulse with a carrier-envelope phase stabilization [6] and a wiggler with one or two periods upstream of the FEL undulator, and thus obtain energy modulation of electrons as shown in Fig.1. Two special regions are seen there: one with the largest energy offset and one with the largest energy chirp. The uniqueness of each region can be explored in the FEL in a number of ways [7, 8, 9, 10, 11, 12, 13, 14], each favoring lasing either by electron group (1) or group (2), and damping lasing of the remaining electrons. For example, a tapered undulator producing  $z$  dependence of the undulator parameter  $K$  defined in [10] as

$$\frac{d \ln K}{dz} = -\frac{\lambda_x}{\lambda_u} \frac{1+K^2/2}{K^2/2} \frac{d \ln \gamma}{cdt}, \quad (3)$$

with  $\lambda_x$  the x-ray wave length and  $\lambda_u$  the undulator period, will only support lasing of electrons with the energy chirp  $d\gamma/dt$  from region 2. Depending on the actual technique, FEL type, and x-ray photon frequency, the minimum duration of x-ray pulses can vary from a few femtoseconds to a few tens of attoseconds. However, the time

bandwidth product of these pulses is close to the Fourier transform limit in most cases.

The most attractive feature of all techniques involving the laser is synchronization of the ultrashort x-ray pulse to a modulating laser pulse. Moreover, two or more successive laser pulses can produce two or more electron beam modulations in different sections of the electron bunch and, therefore, create two or more ultrashort x-ray pulses with a precise time delay between them. It is even possible to have two-color ultrashort x-ray pulses [14].

## References

- [1] J. Feikes et al, EPAC 04, 1954
- [2] S. Reiche et al, NIM A593 (2008) 39
- [3] R.W. Schoenlein et al, Science 287 (2000) 2237
- [4] A.A. Zholents et al, NIM A425 (1999) 385
- [5] P. Emma et al, PRL 92, 74801 (2004)
- [6] F. Krausz, M. Ivanov, RMP 81 (2009) 163
- [7] A.A. Zholents, W.M. Fawley, PRL 92, 224801 (2004)
- [8] E.L. Saldin, E.A. Schneidmiller, M.V. Yurkov, Opt. Commun. 239 (2004) 161
- [9] A.A. Zholents, G. Penn, PRST-AB 8, 050704 (2005)
- [10] E.L. Saldin, E.A. Schneidmiller, M.V. Yurkov, PRST-AB 9, 050702 (2006)
- [11] A.A. Zholents, M.S. Zolotorev, NJP 10, 025005 (2008)
- [12] Y. Ding et al. PRST-AB 12 (2009) 060703
- [13] D. Xiang, Z. Huang, G. Stupakov, PRST-AB 12, 060701 (2009)
- [14] A.A. Zholents, G. Penn, NIM A612 (2010) 254

### 3.1.10 Compton/Thomson Sources

*G.A. Krafft, TJNAF*

When a relativistic electron beam passes through a high-field laser beam, the electrons radiate intense and highly collimated electromagnetic radiation by the Compton/Thomson scattering process. Due to relativistic upshifting of the photons and the relativistic Doppler effect, highly energetic polarized photons are radiated along the electron beam motion. By this means effective x-ray and gamma ray sources can be built [1]. High energy Compton scattering is discussed in Sec.3.3.4. The physics and quantitative discussion of the scattered radiation is largely the same as in undulators (Sec.3.1.5). In Fig.1, the  $z$ -axis is aligned with the relativistic electron beam (relativistic factors  $\beta$  and  $\gamma$ ), the wave vector of the

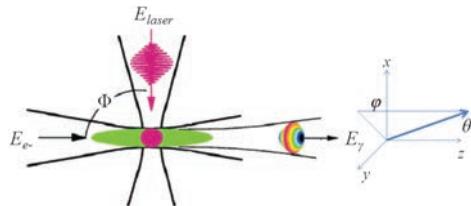


Figure 1: Scattering geometry and angle definitions.

Table 1: Energy Reach of Scenarios.

| Scenario               | Maximum Energy                           |
|------------------------|--|
| Backscattering         | $4\gamma^2 E_{\text{laser}}$             |
| Sidescattering         | $2\gamma^2 E_{\text{laser}}$             |
| Small Angle Scattering | $(1 + \gamma^2 \Phi^2) E_{\text{laser}}$ |

incident laser beam is in the  $x$ - $z$  plane at angle  $\Phi$ , and the laser polarized linearly in  $y$ -direction.  $\theta$  and  $\phi$  are spherical polar angles for the scattered radiation. Definitions: Backscattering  $\Phi = \pi$ , Sidescattering  $\Phi = \pi/2$ , Small Angle Scattering  $\Phi \ll 1$ .

The incident photon energy in the beam frame is  $E_{\text{laser}}\gamma(1 - \beta \cos \Phi)$  where  $E_{\text{laser}}$  is the incident photon energy in the lab. In terms of the angles in Fig.1, the scattered energy  $E_\gamma$  is

$$E_\gamma(\theta, \phi) = \frac{E_{\text{laser}}(1 - \beta \cos \Phi)}{1 - \beta \cos \theta + E_{\text{laser}}(1 - \cos \Delta\Theta)/E_e}$$

where  $\Delta\Theta$  is the angle between the incident and scattered photons satisfying  $\cos \Delta\Theta = \cos(\Phi) \cos(\theta) - \sin \Phi \sin \theta \cos \phi$ , and  $E_e = \gamma mc^2$  is the total energy of the electrons. The Thomson limit, implying the neglect of electron recoil, is [2]

$$E_\gamma(\theta, \phi) = \frac{E_{\text{laser}}(1 - \beta \cos \Phi)}{1 - \beta \cos \theta}$$

There are non-linear corrections at high laser intensities discussed separately.

A plane wave expansion for the incident photon beam in terms of normalized vector potential ( $a = eA/mc$ ) is

$$\mathbf{a}(\mathbf{x}, t) = \Re\{a(\omega t - \mathbf{k} \cdot \mathbf{x})e^{i(\omega t - \mathbf{k} \cdot \mathbf{x})}\}\hat{\mathbf{y}}$$

The peak value of the normalized vector potential is generally denoted  $a_0$ . This field strength parameter is analogous to  $K$  for undulators. A plane wave formulation implies the neglect of hour-glass effects in the description of the colliding photons.

### Sec.3.1: SYNCHROTRON RADIATION

The spectral energy density of the output pulse per electron is [3]

$$\frac{d^2U_\gamma}{d\omega d\Omega} = \frac{r_e^2 \epsilon_0}{2\pi c} \left| \tilde{E} \left[ \frac{\omega(1-\beta \cos \theta)}{c(1-\beta \cos \Phi)} \right] \right|^2 \times \frac{\cos^2 \phi (1-\beta \cos \theta)^2 + \sin^2 \phi (\cos \theta - \beta)^2}{\gamma^2 (1-\beta \cos \theta)^4},$$

where  $\tilde{E}$  is the spatial Fourier transform of the transverse electric field as a function of the longitudinal variable of the incident laser evaluated in the lab frame ( $E = -\partial A / \partial t$ ), and the notation means evaluate the transform at the Doppler shifted wave number  $\omega(1-\beta \cos \theta)/c(1-\beta \cos \Phi)$ .

The number of photons scattered per electron in a flat incident pulse is ( $a_0 \ll 1$ )

$$N_{\text{per } e-} = \frac{2\pi\alpha N_\lambda a_0^2}{3},$$

where  $\alpha$  is the fine-structure constant and  $N_\lambda$  is the number of wavelengths in the incident pulse.

#### 3.1.10.1 Luminosity description

Compton sources are an example of a photon-lepton collider. A description in terms of “luminosity” is appropriate. Results in this section are specific to the Thomson Backscatter case.

In a single collision, the total number of scattered photons is Lorentz invariant, as is  $\sigma$ , the total scattering cross section in the rest frame of the electrons [3, 4]. The lepton current 4-vector is  $\Phi_e^\mu = J^\mu/e$ . The incident laser photons are described by the 4-vector  $\Phi_p^\mu = (n_p, n_p \hat{k}_p)$ . The number of scattered photons  $N_\gamma$  is proportional to both  $n_p$  and  $n_e$ , and  $d^4x$  is Lorentz invariant. Consequently,  $N_\gamma$  must be

$$N_\gamma = \int \Phi_{p\mu} \Phi_e^\mu \sigma d^4x = \int (1 - \beta_p \cdot \beta_e) n_p(x, y, z, t) n_e(x, y, z, t) \sigma d^4x.$$

When the hourglass effect is negligible the distributions are separable,  $n_p(x, y, z, t) = n_p(x, y) I_p(z + ct)$ ,  $n_e(x, y, z, t) = n_e(x, y) I_e(z - \beta ct)$ , where  $I_p, I_e$  are unit normalized, and

$$N_\gamma = \int n_p(x, y) n_e(x, y) \sigma_T dx dy,$$

where the total Thomson scattering cross section is  $\sigma_T = 8\pi r_e^2/3$ . For overlapped Gaussians of rms size  $\sigma_e$  and  $\sigma_p$ ,

$$N_\gamma = \frac{N_p N_e}{2\pi \sqrt{\sigma_{ex}^2 + \sigma_{px}^2} \sqrt{\sigma_{ey}^2 + \sigma_{py}^2}} \sigma_T$$

where  $N_p$  is the number of incident photons and  $N_e$  is the number of electron scatterers. The luminosity is

$$L = f \frac{N_p N_e}{2\pi \sqrt{\sigma_{ex}^2 + \sigma_{px}^2} \sqrt{\sigma_{ey}^2 + \sigma_{py}^2}},$$

where  $f$  is the collision repetition rate. A related formula is

$$U_\gamma = \gamma^2 (1+\beta) \frac{N_e \sigma_T}{2\pi \sqrt{\sigma_{ex}^2 + \sigma_{px}^2} \sqrt{\sigma_{ey}^2 + \sigma_{py}^2}} U_{\text{laser}}$$

where  $U_\gamma$  is the total scattered energy per pulse and  $U_{\text{laser}}$  is the total energy in the incident pulse.

#### 3.1.10.2 Nonlinear scattering

As  $a_0$  approaches one, frequency red-shifting and harmonic generation become prominent [5]. For a longitudinally flat illuminating laser, computations and results are very similar to those in standard undulator theory.

E.g., a modified resonance condition

$$E_\gamma \approx \frac{4\gamma^2 E_{\text{laser}}}{1 + a_0^2/2}$$

applies for linear polarization. As in undulators,  $a_0^2/2 \rightarrow a_0^2$  in this formula for a circularly polarized incident laser.

For linear polarization, odd harmonics are generated. The strengths of the harmonic lines are quantified by “Bessel function factors”. These factors have the same functional form as in undulators, but the relativistic invariants in the arguments of the functions are modified. For example, the flux factor into the  $n$ th harmonic ( $n$  odd) is [6]

$$F_n(a) = \frac{n^2 a_0^2}{(1 + a_0^2/2)^2} \left\{ J_{(n-1)/2} \left( \frac{n a_0^2}{4(1 + a_0^2/2)} \right) - J_{(n+1)/2} \left( \frac{n a_0^2}{4(1 + a_0^2/2)} \right) \right\},$$

where  $J$  denotes the Bessel function of integer order. At low  $a_0$ , the strength of the higher harmonics grows with much higher powers of the field strength than the fundamental. As  $a_0$  increases

beyond 1, most of the emission is radiated in the harmonics.

For non-flat illumination pulses the situation is more complicated. Because of electron slow-down and red-shifting, the spectrum emitted becomes broad, exhibits interference structures, and is generally maximum at the maximum red shift within the pulse [7].

## References

- [1] P. Sprangle et al, JAP 72 (1992) 5032
- [2] F. V. Hartemann, et al, PRST-AB 8, 100702 (2005)
- [3] G.A. Krafft, PAC 97, p.739
- [4] W.J. Brown, F.V. Hartemann, PRST-AB 7, 060703 (2004)
- [5] E. Esarey, S.K. Ride, P. Sprangle, PR E48 (1993) 3003
- [6] K.-J. Kim, AIP Conf. Proc. 184 (1989) 565
- [7] G.A. Krafft, PRL 92, 204802 (2004)

### 3.1.11 “Short” Magnet and Edge Radiation

*H. Wiedemann, Stanford U./SSRL*

The source length of synchrotron radiation from long uniform magnets is  $\sim \pm \rho/\gamma$  setting an upper limit for the photon energy spectrum  $\hbar\omega_c \approx 4\pi\hbar c\gamma^3/\rho$ . If a magnet is shorter than this source length ( $L < \rho/\gamma$ ), the radiation pulse is also shorter and the spectrum extends to higher frequencies of about  $\omega_{sh} \approx 4\pi c\gamma^2/L$ . The spatial distribution of the photon flux into a band width  $d\omega/\omega$  is [1, 2]

$$\frac{dN_{ph}(\omega)}{d\Omega} = \frac{8\epsilon_0 c^3 r_e^2}{\pi\hbar} \frac{I}{e} \frac{d\omega}{\omega} \gamma^6 f^2 |\tilde{B}(\omega)|^2 \quad (1)$$

with

$$f^2 = \frac{(1 + \gamma^2\theta^2)^2 \sin^2 \varphi + (1 - \gamma^2\theta^2)^2 \cos^2 \varphi}{(1 + \gamma^2\theta^2)^6}$$

where the two terms in the denominator indicate the  $\sigma$ - and  $\pi$ - mode polarization respectively,  $\theta$  the angle between observation ( $n$ ) angle with respect to the beam axis ( $v$ ) and  $\varphi$  the angle between plane ( $n, v$ ) and plane including acceleration and beam ( $a, v$ ),  $\tilde{B}(\omega)$  is the Fourier transform of the magnetic field  $B(t)$ . In case of a Lorentzian field distribution  $B(z) = B_0/(1 + 4z^2/L^2)$ , Eq.(1) becomes

$$\frac{dN_{ph}(\omega)}{d\Omega} = \frac{\pi\epsilon_0 c r_e^2}{2\hbar} \frac{I}{e} \frac{d\omega}{\omega} \gamma^2 B_0^2 L^2 f^2 \times (1 + \gamma^2\theta^2)^2 \exp \left[ -\frac{L(1 + \gamma^2\theta^2)}{2\gamma^2 c} \omega \right]$$

Radiation with fundamentally different characteristics is being emitted by relativistic electrons at the entrance and exit edge of otherwise uniform bending magnets. If  $\vartheta$  and  $\psi$  are the horizontal and vertical observation angles with respect to the beam axis and horizontal being the deflecting plane, we get for the spatial distribution of the photon flux [3]:

$$\frac{dN_{ph}(\omega)}{d\Omega} = \frac{\alpha}{\pi^2} \gamma^2 \frac{I}{e} \frac{d\omega}{\omega} (F_\sigma \hat{e}_\sigma + F_\pi \hat{e}_\pi)^2$$

where

$$F_\sigma = \frac{i\pi}{\lambda\gamma} \int_{-\infty}^{\infty} (\theta - \beta_x) \exp(i\Phi) dz$$

$$F_\pi = -\frac{i\pi\psi}{\lambda\gamma} \int_{-\infty}^{\infty} \exp(i\Phi) dz$$

$$\Phi = \frac{\pi}{\lambda} [(\theta^2 + \psi^2 + 1/\gamma^2)z - 2\theta \int_0^z \beta_x dz' + \int_0^z \beta_x^2 dz']$$

$\beta_x = v_x/c$  the transverse particle velocity,

$$\beta_x(z) = \frac{1}{\rho} \int_0^s f(z') dz' + \beta_x(0) \quad (2)$$

with  $\rho = -m_e c^2 \gamma / (e B_0)$  and  $f(z) = B(z)/B_0$ , assuming that the particles oscillate in the deflecting/horizontal plane only.

There is no radiation in the forward direction. The intensity of the  $\sigma$ -mode increases with  $\theta$  reaching a maximum at  $\theta = 1/\gamma$ . The  $\pi$ -mode radiation peaks at  $\psi = 1/\gamma$ . For very long wavelengths the photon flux levels off at

$$\frac{dN_{ph}(\omega)}{d\Omega} \Big|_{\theta^2 + \psi^2 = 1/\gamma^2}^{\lambda \rightarrow \infty} \rightarrow \frac{\alpha \gamma^2}{4\pi^2} \frac{I}{e} \quad (3)$$

## References

- [1] R. Coisson, PR A20 (1979) 524
- [2] R. Bossart et al, NIM 164 (1979) 375
- [3] O.V. Chubar, N.V. Smolyakov, PAC 93, 1626

### 3.1.12 Beam Solid-Target Photon Physics

*K. Ispiryan, Yerevan Physics Institute*

This section reviews Cherenkov radiation, bremsstrahlung, coherent bremsstrahlung, channeling radiation, parametric x-ray radiation, and Smith-Purcell radiation [1, 2, 3, 4, 5, 6].

### Sec.3.1: SYNCHROTRON RADIATION

**Cherenkov radiation (CR):** Electromagnetic radiation [1] produced by particles with charge  $z$  passing through a medium with constant velocity  $\mathbf{v}$  exceeding the phase velocity  $v/n$  of light in the medium with refractive index  $n(\omega)$ , i.e. CR has a threshold  $\beta_{\text{thr}} = v_{\text{thr}}/c = 1/n$ . With the help of Weizsäcker-Williams method of pseudophotons accompanying the particles one can interpret the production of CR photons as W-W photons separated from the charge due to their slowness. CR has been discovered in 1934 [7], while the QED theory has been developed in [8]. From simple Huygens principles, the CR photons are emitted under angle

$$\cos \theta = 1/\beta n \quad (1)$$

with spectral distribution of number of photons per unit length equal to

$$\frac{d^2N_{\text{CR}}}{dld(\hbar\omega)} = \frac{z^2\alpha}{\hbar c} \sin^2 \theta, \quad (2)$$

where  $\alpha = e^2/hc = 1/137$ ,  $hc = 1.9733 \times 10^{-5}$  eVcm. CR spectrum is extended in certain wide or narrow spectral regions from radio up to X-ray frequencies where  $n(\omega) > 1$ . CR has found wide application in particle detection and identification as well as in some advanced particle acceleration methods.

**Bremsstrahlung (B)** Electromagnetic radiation [2, 3] produced when the velocity of charged particles (mainly of light electrons) is changed due to scattering or deceleration in the field of nuclei and electrons. Quantum electrodynamical (QED) theory of B has been developed in [9]. One can interpret and even obtain some results considering the Compton scattering of pseudophotons on electrons and nuclei. When the electron energy is higher than the so called critical energy,  $E_{\text{cr}} \approx 800 \text{ MeV}/(Z+1)$ , radiation energy losses due to B are higher than ionization energy losses and increase proportionally to  $\exp(x/X_0)$  with  $x$  the target thickness and  $X_0$  the so called radiation length, which at  $\gamma = E/mc^2 \gg 1$  is equal to

$$X_0 = \frac{A716.4}{Z(Z+1)\ln(287/\sqrt{Z})}, \quad (3)$$

where  $A$  and  $Z$  are the mass and atomic number of the radiator nuclei. At electron energies  $mc^2 \ll E \leq 50 \text{ MeV}$  one can estimate the angular-spectral distribution classically:

$$\frac{d^2\sigma}{d\omega d\Omega} = \frac{8Z^2 r_0^2 \gamma^2}{137\pi\omega} \ln\left(\frac{233}{Z^{1/3}}\right) \frac{1 + \gamma^4 \theta^4}{(1 + \gamma^2 \theta^2)^4}.$$

As it is seen the dipole like  $\sim \cos^2 \theta$  angular distribution of B valid for  $E \ll mc^2$  becomes sharply forward directed with maximum at  $\theta \sim 1/\gamma$ , while the spectral distribution is  $\sim 1/\omega$ .

At higher energies the spectral distribution of B is given by the following QED formula (also relevant for other types of radiation)

$$\frac{d\sigma_B}{d\omega} = 4\alpha Z^2 r_0^2 \frac{1}{\omega} \left[ 1 + (1-x)^2 \psi_1(\delta) - \frac{2}{3} (1-x) \psi_2(\delta) \right], \quad (4)$$

where  $x = \hbar\omega/E$ ,  $r_0 = e^2/mc^2$  the classical electron radius,

$$\delta = \frac{\hbar\omega mc^2}{2E(E - \hbar\omega)} \frac{2\pi}{\lambda_C} = \frac{1}{2} \frac{2\pi}{\lambda_C} \frac{1}{\gamma} \frac{x}{1-x} \quad (5)$$

the minimal value of the longitudinal component  $q_{||}$  of the momentum transfer  $q$ ,  $\lambda_C = (h/mc)$  the electron Compton wavelength,

$$\psi_1(\delta) = 1 + \int_{\delta}^{mc} f_{\text{scr}}(q)(q-\delta)^2 q dq, \quad (6)$$

$$\psi_2(\delta) = \frac{5}{6} + \int_{\delta}^{mc} (q^3 - 6\delta^2 q \ln \frac{q}{\delta} + 3\delta^2 q - 4\delta^3) f_{\text{scr}}(q) dq, \quad (7)$$

and  $f_{\text{scr}}(q)$  is a factor taking into account the screening by the atomic electrons when the screening parameter  $\zeta = 100/[Z^{1/3}\gamma(1/x-1)]$  is not much greater than unity. In Moliere approximation  $f_{\text{scr}}(q)$  is given by

$$f_{\text{scr}}(q) = \left[ \sum_1^3 \frac{\alpha_i}{\beta_i^2 + q^2} \right], \quad (8)$$

where  $\alpha_1 = 0.1$ ;  $\alpha_2 = 0.55$ ;  $\alpha_3 = 0.35$ ;  $\beta_1 = (Z^{1/3} b_i/121)$ ;  $b_1 = 6$ ;  $b_2 = 1.2$ ;  $b_3 = 0.3$ . In the case of complete screening  $\psi_1 = \ln 183Z^{-1/3}$  and  $\psi_2 = \ln 183Z^{-1/3} - 1/6$ . The angular distribution of B is narrow,  $\theta \sim 1/\gamma$ .

In certain cases it is necessary to take into account 1) The medium polarization or Ter-Mikaelian effect resulting in suppression of the spectral distribution at low photon energies  $\hbar\omega < \hbar\omega_p\gamma$  where  $\omega_p$  is the plasma frequency and 2) The Landau-Pomeranchuk-Migdal (LPM) effect at very high electron energies resulting in suppression of the spectra at low photon energies  $\hbar\omega < E^2/E_{\text{LPM}}$ , where  $E_{\text{LPM}} = \alpha m^2 X_0 / 8\pi \hbar c = 3.8 \times 10^{12} \text{ eV } X_0 [\text{cm}]$ .

**Coherent Bremsstrahlung (CB):** CB [2, 4] is produced when charged particles pass through crystal under small angle (but larger than the channeling critical angles, see below) to the crystallographic axes and/or planes in crystals and when in some frequency regions the intensity of B is enhanced due to interference of coherent photons. For history of initial CB theory and experiments see [2, 4]. The production of CB can be considered as follows [2, 4]: CB photons are formed over a longitudinal distance called formation zone,

$$Z_{\text{coh}} = \frac{1}{\delta} = 2\lambda_C \gamma \frac{1-x}{x}. \quad (9)$$

For values of  $\gamma$  sufficiently high,  $Z_{\text{coh}}$  becomes larger than the periodic distances,  $d$ , between the atoms in the axes or between the planes. Therefore, the production of B photons in certain energy regions on  $\sim N_{\text{coh}} \sim Z_{\text{coh}}/d$  nuclei takes place coherently, and one can expect enhancement approximately  $\sim N_{\text{coh}}^2$  times.

Now the radiation cross section is the sum of coherent and incoherent parts. As the calculations show [2, 4] both parts are given by expressions of type (4) with

$$\psi_1^{\text{coh}}(\delta) = \frac{(2\pi)^2}{Na^3} \sum_g |S(g)|^2 e^{-2W} f_{\text{scr}}(g) \frac{\delta g_{\perp}^2}{g_{||}^2}, \quad (10)$$

$$\psi_2^{\text{coh}}(\delta) = 6 \frac{(2\pi)^2}{Na^3} \sum_g |S(g)|^2 e^{-2W} f_{\text{scr}}(g) \frac{\delta^2 g_{\perp}^2 (g_{\perp} - \delta)}{g_{||}^4}. \quad (11)$$

In (10) and (11)  $N$  is the number of atoms per unit cell,  $a$  the crystal constant,  $g_{||}$  and  $g_{\perp}$  are the longitudinal and transverse components of the reciprocal lattice vector  $\mathbf{g}$  with  $g = 2\pi/d_{hkl}$  and  $d_{hkl}$  the interplanar spacing defined by the Miller indices  $h, k, l$ , and the summation is carried out over all  $\mathbf{g}$  with  $g_{||} > \delta$ .

$$|S(g)|^2 = \left| \sum_{j=1}^N e^{i\vec{g}\vec{r}_j} \right|^2, \quad (12)$$

is the structure factor and the factor  $e^{-2W}$  the Debye-Waller factor, where  $W = 0.5 < u_1^2 > g^2$ , with  $< u_1^2 >$  the mean square of 1D thermal amplitude of the crystal atoms. For the incoherent part of the cross section the expressions under the integrals (6) and (7) are multiplied by a factor

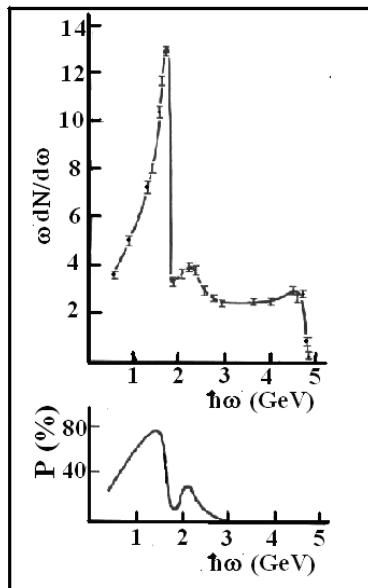


Figure 1: The spectral distribution of the intensity of CB measured at YerPhi using 4.5 GeV electron and 100  $\mu\text{m}$  thick diamond (points with errors) and calculated (solid curve, top figure) and the calculated polarization dependence on photon energy (solid curve, bottom figure) (by courtesy of R.O. Avakian).

$(1 - e^{-2W})$  which decreases the incoherent cross section by  $\sim 10\%$  compared with (4).

CB photons are polarized, and degree of linear polarization is equal to

$$P = \frac{d\sigma_{||} - d\sigma_{\perp}}{d\sigma_{||} + d\sigma_{\perp}} = \frac{2uv^2}{1 + u^2 - 4uv(1-v)}, \quad (13)$$

with  $u = 1 - x$ ,  $v = \delta/q_{||}$ ,  $q_{||}$  the longitudinal component of the transfer momentum. From (13) the polarization is maximal at  $v = 1$ . Usually the polarization is calculated from the measured spectral distribution because the direct measurement of polarization is much more difficult. Figure 1 shows the dependence of intensity and polarization on the energy of the CB photons produced by 4.5 GeV electrons passing through a 100  $\mu\text{m}$  thick diamond crystal oriented for maximal polarization.

CB photons have found wide application in photoproduction experiments.

### Sec.3.1: SYNCHROTRON RADIATION

**Channeling Radiation (ChR)** Planar and axial ChR [4, 11] is produced by channeled particles entering into crystal under angles smaller than the so called critical or Lindhard angles,  $\theta_{cr} = (2U/E)^{1/2}$ , with respect to crystallographic planes and axes with averaged potential depth U, respectively. When the electron energy is higher than a few hundred MeV, and the number of transverse energy levels of the particles in the potential well is much larger than one, one can consider ChR classically as radiation of a moving dipole. At lower energies quantum approach is necessary, and ChR is produced as a result of transitions between a finite number of transversal energy levels of electrons in crystal. Apparently ChR has been observed for the first time in the works [10], while the corresponding theory has been developed by many authors gradually up to its transparent interpretation (see [11]).

In dipole approximation, applicable roughly when  $\theta_{cr} \gamma < 1$ , the spectral distribution of ChR like the one for undulator radiation consists of many harmonics with numbers  $n = 1, 2, 3, \dots$  and for a parabolic potential  $U(x) = U_0 x^2$  is given by [11]

$$\frac{dI_{ChR}}{d\omega} = \frac{e^2}{c} \sum_{n=1}^{\infty} \frac{|\dot{\beta}_{\perp n}|^2}{\Omega_n^2} \omega \left[ 1 - 2 \frac{\omega}{\omega_n} + 2 \frac{\omega^2}{\omega_n^2} \right] \quad (14)$$

where  $\omega_n = 2\Omega_n \gamma^2$ ,  $\Omega_n = n(2U_0/m)^{1/2}$  the oscillation frequency,  $\dot{\beta}_{\perp n}$  the transverse Fourier component of the acceleration (for examples see [11]). In the “dipole” case the first harmonic is dominant. For undulator radiation the particle oscillation period is given by the undulator period, and  $\omega_{max} \sim \gamma^2$ , while for ChR the particle oscillation period depends on particle energy and  $\omega_{max} \sim \gamma^{2/3}$ . As in the case of undulator and many other types of radiation ChR photons are emitted under angles  $\sim 1/\gamma$ , and there is a relation between the frequency and the emission angle.

Figure 2 shows the spectral distribution of the number of ChR photons produced by 6 GeV positrons channeled in a 100  $\mu m$  thick diamond.

At electron energies lower than a few hundreds MeV ChR spectral-angular distribution is calculated quantum mechanically by solving the Schrödinger equations to find the energy levels, the population and transition probabilities over all the crystal thickness. Also multiple scattering, dechanneling, polarization of ChR etc. must be taken into account. The characteristics

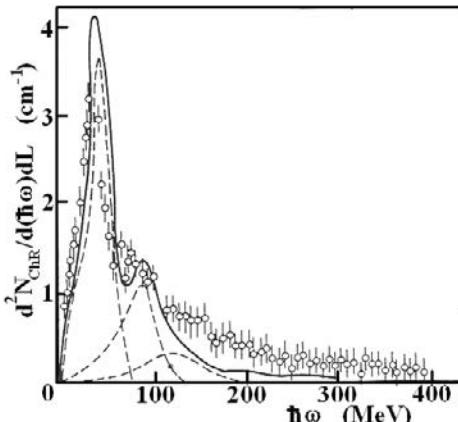


Figure 2: The spectral distribution of ChR of positrons in 100  $\mu m$  thick diamond. The circles are the results of measurements at SLAC, the curves are the results of calculation for the first 3 harmonics (dashed curves) and for the sum (solid curve) (by courtesy of R.O. Avakian).

of ChR are close to those of transition radiation for electron energies of a few tens of MeV. Like the transition radiation ChR has found wide application.

**Parametric X-ray Radiation (PXR):** PXR [4, 6] is produced when charged particles pass the crystallographic planes. It can be considered as Bragg diffraction of pseudophotons. Differently from most of the other types of radiation, PXR is monochromatic and emitted under large angles  $\theta \gg 1/\gamma$ , close to Bragg angles. It was predicted as a special case of resonance transition radiation in 1969 in the Russian publication of the monograph [2] using the perturbation theory of the classical electrodynamics and was first observed in [12]. Like the theory of Bragg diffraction the PXR theory was developed in kinematical (correct for thin crystals when multiple reflections from the crystallographic planes can be neglected) and dynamical approaches [13].

According to the kinematical approximation of PXR (the more general dynamical and quantum mechanical theories of PXR give no essential corrections) the number of photons  $N_{PXR}$  produced by an electron in a crystal with thickness  $L$  and absorption length  $L_{abs}$ , per unit angles  $\theta_x$  and  $\theta_y$  — the parallel and perpendicular components of the PXR emission angle  $\theta = \sqrt{\theta_x^2 + \theta_y^2}$  with

respect to the Bragg reflection direction — is

$$\frac{d^2N_{\text{PXR}}}{d\theta_x d\theta_y} = \frac{\alpha}{4\pi} \omega_B F_{\text{geom}}(\hat{\sigma}, \hat{v}, \hat{k}) \quad (15)$$

$$\frac{|\chi_g(\omega_B)|^2 \theta_x^2 \cos^2(2\theta_B) + \theta_y^2}{\sin^2 \theta_B \left( \theta_x^2 + \theta_y^2 + \theta_{ph}^2 \right)^2},$$

where  $\omega_B$  and  $\theta_B$  are the Bragg frequency and angle for the given crystal orientation,

$$F_{\text{geom}}(\hat{g}, \hat{v}, \hat{k}) = L_{\text{abs}} \frac{|\hat{\sigma} \hat{k}|}{|\hat{\sigma} \hat{v}|} \quad (16)$$

$$\left[ 1 - \exp \left( -\frac{L}{L_{\text{abs}} |\hat{\sigma} \hat{k}|} \right) \right]$$

is a geometrical factor;  $\hat{\sigma}$ ,  $\hat{v}$  and  $\hat{k}$  are unit vectors in directions perpendicular to the surface, of the particle particle  $\mathbf{v}$ , and of the emitted photon, respectively;

$$X_g(\omega) = S(g) \exp(-W) \frac{\omega_p^2 F(g)}{\omega^2 Z} \quad (17)$$

is the coefficient of the Fourier expansion of the dielectric susceptibility;  $F(g)$  the atomic scattering factor; and

$$\theta_{ph}^2 = \frac{1}{\gamma^2} + \frac{\omega_p^2}{\omega^2}. \quad (18)$$

In Eq.(15) the frequency is related to the PXR angle through

$$\omega = \frac{c |\hat{v} \hat{g}|}{1/\beta - \sqrt{\epsilon} \hat{v} \hat{k}}, \quad (19)$$

which is close to the Bragg formula for X-ray diffraction. In the last expression  $\epsilon = 1 - (\omega_p/\omega)^2$  is the dielectric constant.

For various crystals and geometries the total number  $N_{\text{PXR}}$  of emitted PXR photons per electron increases with  $\gamma$  up to  $\gamma_{\text{cr}} \sim \omega_B/\omega_p \sim 200$ , saturating at  $N_{\text{PXR}} \sim 10^{-5} - 10^{-7}$ . Beginning from  $\gamma_{\text{cr}}$  the contribution of the diffracted transition radiation (DTR) photons with almost the same frequency and emission angle slightly exceeds the pure PXR yield.

**Smith-Purcell Radiation (SPR)** SPR [6] is the coherent diffraction radiation produced by charged particles flying over a grating with period  $d$  at distance  $h$  of the order  $\gamma \beta \lambda_{\text{SP}}/(4\pi)$ . It was discovered and studied in [14] in which using Huygens principles it has been shown that for the  $n$ -th harmonic

$$\lambda_{\text{SP}} = \frac{d}{n} \left( \frac{1}{\beta} - \cos \theta \right)$$

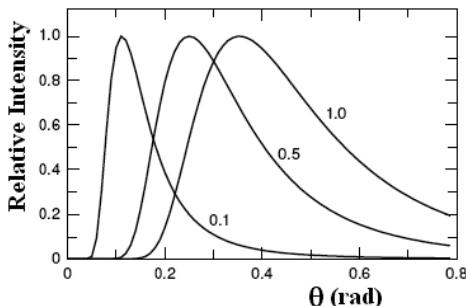


Figure 3: Relative intensity of SPR vs  $\theta$  for various values of  $h/d$  indicated next to the curves.

Many models for analytical and numerical calculations for the angular distribution of the intensity  $d^2I/dzd\Omega$  of SPR per unit length of the grating have been developed [6]. At relativistic energies [15]

$$\frac{d^2 I_{\text{SP}}}{dz d\Omega} = \frac{4e^2}{\pi d^2} \sin^2 \theta \exp \left[ -\frac{4\pi h/d}{\gamma (1 - \beta \cos \theta)} \right]$$

$$\frac{1}{(1 - \beta \cos \theta)^3}$$

Figure 3 shows SPR angular distribution.

At present SPR has been experimentally studied in the region from millimeter up to optical wavelengths using electrons with  $E = (0.02 - 855)$  MeV. SPR can be used for constructing tunable compact sources (also SPR FELs) and for diagnostics of particle beams.

## References

- [1] J.D. Jackson, Classical Electrodynamics, 3rd edition, J. Wiley&Sons, New York, 1999
- [2] M.L. Ter-Mikaelian, The Influence of the Medium on High Energy Processes at High Energies, Academy of Science of Armenia, Yerevan 1969; High Energy Electromagnetic Processes in Condensed Media, Wiley Interscience, New York, 1972
- [3] V.B. Berestetski, E.M. Lifshitz, L.P. Pitaevski, Kvantovaya Elektrodynamika, Moscow, Nauka, 1989 (in Russian)
- [4] P. Rullhusen, X. Artru, P. Dhez, Novel Radiation Sources Using Relativistic Electrons, World Scientific, Singapore, 1998
- [5] H. Wiedemann, Synchrotron Radiation, Springer, Berlin, 2003
- [6] A.P. Potylitsin, Electromagnetic Radiation of Electrons in Periodic Structures, Springer, Berlin-Heidelberg, 2011

### Sec.3.2: IMPEDANCES AND WAKE FUNCTIONS

- [7] P.A. Cherenkov, Doklady Akad.Nauk SSSR, v.2 (1934) 451
- [8] I.E. Tamm, I.M. Frank, Doklady Akad.Nauk SSSR, v.14 (1937) 107
- [9] H. Bethe, W. Heitler, Proc.Roy.Soc. 146 (1934) 83
- [10] R.L. Walker et al, PRL 25 (1970) 5; PR A11 (1975) 736
- [11] M.A. Kumakhov, Izluchenie Kanalirovannikh Chastits v Kristallakh, Energoizdat, Moscow, 1986 (In Russian)
- [12] S.A. Vorobev et al, Zh.Eksp.Teor.Fiz. 41 (1985) 3
- [13] V.G. Barishevski, I.D. Feranchuk, A.P. Ulyanov, Parametric X-Ray Radiation in Crystals: Theory, Experiments and Application, Springer Tracts in Modern Physics, Heidelberg, 2005
- [14] S.J. Smith, E.M. Purcell, PR 92 (1953) 1069
- [15] J.E. Walsh, K. Woods, S. Yeager, NIM A341 (1994) 277

## 3.2 IMPEDANCES AND WAKE FUNCTIONS

### 3.2.1 Definitions and Properties of Impedances and Wake Functions T. Suzuki, KEK

#### Driving current

##### Axially symmetric case [1]

[coordinates  $(r, \theta, s)$ ] Charge  $q$ , offset  $r = a$ ,  $\theta = 0$ , velocity  $\vec{v}$  in  $s$ -direction.

$$\begin{aligned} \rho &= \frac{q}{a} \delta(r-a) \delta(\theta) \delta(s-vt) \\ &= \sum_{m=0}^{\infty} \frac{I_m \cos m\theta}{\pi a^{m+1} (1 + \delta_{m0})} \delta(s-vt) \delta(r-a) \\ \vec{j} &= v\rho\hat{s}, I_m = qa^m \quad (m\text{-th moment}) \end{aligned} \quad (1)$$

##### General cross-section [2]

[coordinates  $(x, y, s)$ ]

$$j_s = qv\delta(x)\delta(y)\delta(s-vt) \quad (2) \quad (\text{longitudinal monopole})$$

$$j_s = qv\delta(y)[\delta(x-x_1) - \delta(x+x_1)]\delta(s-vt) \quad (x_1 \rightarrow 0) \quad (\text{transverse dipole}) \quad (3)$$

**Wake potential** The wake force  $\vec{F} = e(\vec{E} + \vec{v} \times \vec{B})$  is integrated over a period  $L$  (holding  $z = s - vt$  fixed). The wake potential

$$\overline{F} = \int_{-L/2}^{L/2} ds \vec{F} \quad (4)$$

is a function of only  $z = s - vt$ , and satisfies the Panofsky-Wenzel theorem [3],

$$\nabla_{\perp} \overline{F}_{\parallel} = \frac{\partial}{\partial z} \overline{F}_{\perp} \quad (5)$$

Here  $\parallel$  denotes longitudinal and  $\perp$  denotes transverse.

#### Definition of wake functions $W_m, W'_m, W_{\perp}, W_{\parallel}$

$$\begin{aligned} \overline{F}_{\perp} &= -eI_m W_m(z) mr^{m-1} \\ &\times (\hat{r} \cos m\theta - \hat{\theta} \sin m\theta) \\ \overline{F}_{\parallel} &= -eI_m W'_m(z) r^m \cos m\theta \end{aligned} \quad (6)$$

Here a prime denotes  $d/dz$ .  $W_m$  is sometimes called transverse wake function and  $W'_m$  longitudinal wake function. Similarly, for a general cross-section,  $W_{\perp}$  and  $W_{\parallel}$  are defined as

$$\begin{aligned} \overline{F}_{\perp} &= -eq 2x_1 W_{\perp}(z) \hat{x} \quad (\text{transverse dipole}) \\ \overline{F}_{\parallel} &= -eq W_{\parallel}(z) \quad (\text{longitudinal monopole}) \end{aligned} \quad (7)$$

#### Properties of wake functions [1]

- $W_m(z) = 0, W'_m(z) = 0$  for  $z > 0$  (causality).
- $W_m(z) \leq 0, W'_m(z) \geq 0$  for  $z \rightarrow 0^-$ .
- $W_m(0) = 0$  (in most cases, except space charge).
- $W'_m(0) = \frac{1}{2}W'_m(0^-)$  (fundamental theorem of beam loading [4]).
- $W'_m(0^-) \geq |W'_m(z)|$  for all  $z$ .
- $\int_{-\infty}^0 W'_m(z) dz \geq 0$ .

Dimensionalities are

$$[W_m] = [\Omega s^{-1} m^{-2m+1}] = [VC^{-1} m^{-2m+1}]$$

$$[W'_m] = [\Omega s^{-1} m^{-2m}] = [VC^{-1} m^{-2m}]$$

$$[W_{\perp}] = [\Omega s^{-1} m^{-1}] = [VC^{-1} m^{-1}]$$

$$[W_{\parallel}] = [\Omega s^{-1}] = [VC^{-1}]$$

#### Definition of impedances

$$Z_{\parallel m}(\omega) = \int_{-\infty}^{\infty} \frac{dz}{v} e^{-i\omega z/v} W_m(z)$$

$$Z_{\perp m}(\omega) = \frac{i}{v/c} \int \frac{dz}{v} e^{-i\omega z/v} W_m(z) \quad (8)$$

Time dependence of  $e^{-i\omega t}$  is assumed. Some authors use  $e^{i\omega t}(j = -i)$ .  $Z^{\perp}(\omega)$  and  $Z^{\parallel}(\omega)$  are defined in the same way as the Fourier transforms of  $W_{\perp}(z)$  and  $W_{\parallel}(z)$  for a general cross-section.

Dimensionalities are

$$[Z_{\parallel m}] = [\Omega m^{-2m}], \quad [Z_{\perp m}] = [\Omega m^{-2m+1}]$$

### Properties of impedance [1]

- $Z_{\parallel m}(\omega) = \frac{\omega}{c} Z_{\perp m}(\omega)$   
(Panofsky-Wenzel theorem [3]).
  - $\begin{cases} Z_{\parallel m}^*(\omega) = Z_{\parallel m}(-\omega) \\ Z_{\perp m}^*(\omega) = -Z_{\perp m}(-\omega) \end{cases}$   
(reality of wake functions).
  - $\begin{cases} \int_0^\infty d\omega \text{Im}Z_{\perp m}(\omega) = 0 \\ \int_0^\infty d\omega \frac{\text{Im}Z_{\parallel m}(\omega)}{\omega} = 0 \\ \text{Re}Z_{\parallel m}(0) = 0 \end{cases}$   
( $W_m(0) = 0$ , in most cases).
  - $\begin{cases} \text{Re}Z_{\parallel m}(\omega) = \frac{1}{\pi} P.V. \int_{-\infty}^\infty d\omega' \frac{\text{Im}Z_{\parallel m}(\omega')}{\omega' - \omega} \\ \text{Im}Z_{\parallel m}(\omega) = -\frac{1}{\pi} P.V. \int_{-\infty}^\infty d\omega' \frac{\text{Re}Z_{\parallel m}(\omega')}{\omega' - \omega} \end{cases}$   
(causality, Hilbert transform)
- The same expressions apply to  $Z_{\perp m}$ .
- $\begin{cases} \text{Re}Z_{\parallel m}(\omega) \geq 0 \text{ for all } \omega \\ \text{Re}Z_{\perp m}(\omega) > 0 \text{ if } \omega > 0, < 0 \text{ if } \omega < 0 \end{cases}$
  - $Z_{\perp 1} \approx \frac{2c}{b^2 \omega} Z_{\parallel 0}, Z_{\perp m} \approx \frac{2c}{b^{2m} \omega} Z_{\parallel 0}, Z_{\parallel m} \approx \frac{2}{b^{2m}} Z_0^{\parallel}$
- These are approximate expressions relating  $Z_{\perp}$  and  $Z_{\parallel}$ . They are exact for resistive round pipe,  $b$  = vacuum chamber radius (in appropriate freq. range, see Sec.3.2.4).

**Limiting value of wake** For periodic, cylindrically symmetric structures whose closest approach to the axis is  $b$ , the steady state wakes have the properties

$$W'_0(0^-) = \frac{Z_0 c}{\pi b^2}, \quad W'_1(0^-) = \frac{2Z_0 c}{\pi b^4} \quad (9)$$

where  $W'_0, W'_1$ , are respectively the longitudinal, transverse (dipole) wakes per unit length and  $0^-$  signifies a test particle position right behind the exciting charge; with  $W_1(0^-) = 0$  and  $Z_0 = 377 \Omega$ . Eq.(9) is true for a resistive pipe [1], a disk-loaded accelerator structure [5], a pipe with small periodic corrugations [6, 7], and a dielectric tube within a metallic pipe [8]. For a non-round periodic structure the result will be a different constant [e.g. for 2 parallel metallic plates separated by  $2b$ ,  $W_{\parallel 1}(0^-) = \pi Z_0 c / (16b^2)$  [9]], but again a constant dependent on transverse dimensions only and independent of material properties. We see that for short bunches the longitudinal wake approaches a fixed maximum, and the transverse wake approaches zero.

### References

- [1] A.W. Chao, Physics of Collective Beam Instabilities in High Energy Accelerators, Wiley (1993)
- [2] R.L. Gluckstern, J. van Zeijts, B. Zotter, PR E47 (1993) 656
- [3] W.K.H. Panofsky, W.A. Wenzel, RSI 27 (1956) 967
- [4] P.B. Wilson, AIP Proc. 87 (1981) 450
- [5] R. Gluckstern, PR D39 (1989) 2780
- [6] A. Novokhatski, A. Mosnier, PAC 97 (1997) p.1661
- [7] K. Bane, A. Novokhatski, SLAC-AP 117 (1999)
- [8] K.-Y. Ng, PR D42 (1990) 1819
- [9] H. Henke, O. Napoli, EPAC 90 (1990) p.1046

### 3.2.2 Impedance Calculation, Frequency Domain

R.L. Gluckstern, Deceased 2008  
S.S. Kurennoy, LANL

The impedance calculation is reduced to the solution of an electrodynamics problem of finding the fields (voltages, forces) produced in the vacuum chamber by a given beam current. Depending upon the way chosen to formulate and solve the problem one may classify the calculation methods (i) by the initial formulation of the problem — in frequency (excitation by a current harmonic with a given frequency) or in time domain (first calculating wakes, and then finding impedances via Fourier transform); (ii) by the method of solution — analytic, semianalytic, or numerical; (iii) by the approach used — electrotechnical (a chamber element is replaced by an equivalent circuit) or electrodynamic (solving the Maxwell equations).

The convenience of a method depends on the geometry of a chamber element and on the frequency range of interest. Below we briefly list the main calculation methods in the frequency domain mentioning some of their applications and limitations. More detailed analyses and references can be found in specialized reviews and books [1, 2, 3, 4, 5, 6, 7, 8]. Explicit expressions for the impedances of some discontinuities are given in Sec.3.2.4.

**Electrotechnical approach** In this approach an element of the vacuum chamber, typically having a complicated geometry, is replaced by an equivalent circuit, such as a transmission line or  $RLC$ -circuit, with circuit parameters (e.g., its resonance frequency,  $Q$ -value, or line length) determined from heuristic considerations. The approach typically allows one to obtain results in a simple form, but is limited to the frequency range (usually low frequencies) in which the replacement is valid.

### Sec.3.2: IMPEDANCES AND WAKE FUNCTIONS

The broad-band resonator model for the impedance of a vacuum chamber can be regarded as an example of this approach [9]. Another example is the calculation of the impedance of strip-line BPMs [10, 11].

**Electrodynmaic methods** listed below solve the Maxwell equations with a given current to find the fields excited in the chamber. While doing so, proper boundary conditions (BCs) should be imposed on the fields. In addition to the metal BC on the vacuum chamber walls, other conditions may be necessary to “close” the problem: (i) *periodicity conditions*; (ii) “closure” conditions — BCs on imaginary surfaces closing the region of interest such as imaginary metal planes in beam pipes far from the cavity to study its impedance at frequencies below the pipe cut-off; (iii) *radiation conditions* for open nonperiodic structures, such as a cavity coupled to a beam pipe.

**Eigenfunction Method** Since a solution of the inhomogeneous Maxwell equation can be represented by a series in eigenfunctions (EFs), the impedance can be expressed as a formal series. The method allows one to evaluate either resonance impedance values for the lowest modes with frequencies below the chamber cut-off whose EFs can be calculated (usually numerically, in the “closed cavity” approximation, using standard codes, e.g., in [12], or analytically [13]), or the low-frequency impedance in the simplest models where all EFs can be found analytically [14]. For a numerical solution one can apply any frequency-domain eigenvalue solver like Superfish [15] or CST [16].

A modification of the *EF-method* allows one to calculate resonances also at frequencies above the cutoff, but only for *periodic structures* [17]. However, it requires computing numerically the structure dispersion characteristics which is time-consuming, especially for long periods.

**Field matching method** In structures which can be subdivided into a few simple subdomains such that the wave equations in each of them allow the separation of variables, the *field matching method*, also called the *partial domain method*, is applied. Solutions in each subregion are expanded into a complete set of EFs with unknown coefficients. Field matching on the subregion interfaces and BCs lead to an infinite system of linear equations for these coefficients, which usually can be truncated to a finite size.

The matching technique was used to calculate the impedance of a cylindrical chamber with an axisymmetric pill-box cavity with fields matched on the pipe radius: for periodic structures [18], or, with added radiation conditions at infinity, for a single cavity [19]. The matrix equation usually allows an analytic solution in the low-frequency limit and gives approximate expressions for resonances; in a general case, the truncated system of equations is solved numerically. Two codes for periodic pill-box structures are well-known: KN7C [20] for axisymmetric modes, and TRANSVRS [21] for azimuthal ones.

The field matching in transverse planes was used to calculate the impedance of an arbitrary axisymmetric structure with a discontinuity approximated by a sequence of stepwise transitions [22]. The technique is applied naturally to collimators [23] and irises [24]. Besides, the approach can be used for an abrupt step between waveguides of different sizes [25].

**Perturbation methods** For small variations of the transverse dimensions of the chamber *perturbation methods* are applicable. If the radius of an axisymmetric chamber is  $r = b(z) = b(1 + \varepsilon z)$ , where  $\varepsilon = h/2b \ll 1$  is a small parameter, the BC on the perfectly-conducting wall

$$[E_z + b'(z)E_r]_{r=b(z)} = 0, \quad (1)$$

as well as the fields, can be expanded in powers of  $\varepsilon$ . Here  $h$  is the maximal depth of variations, and  $b$  is the mean chamber radius. It leads to a recurrent sequence of equations with BC at  $r = b$ , and to the impedance in the form of a series  $Z = Z^{(0)} + \varepsilon^2 Z^{(2)} + \varepsilon^3 Z^{(3)} + \dots$ , where  $Z^{(0)}$  is the impedance of the smooth chamber and  $Z^{(1)} = 0$ . The technique is known in mathematics as “BC transfer to the regular boundary”. It was applied in the lowest order to calculate the low-frequency impedance and resonances [26]; the higher orders in  $\varepsilon$  were considered in [27].

The  $\varepsilon$ -expansion works when  $b(z)$  is a single-valued function of  $z$  and its derivative  $|b'(z)|$  is limited. An additional restriction applies at low frequencies:  $D \gg h$ , where  $D$  is the structure period [28]; the method fails for bellows corrugations, in which the period  $D$  is comparable with  $h$ .

The impedance of a corrugated vacuum chamber can be calculated at low frequencies [28] and in the resonance region [29] by the **matrix method**. It is applicable to periodic structures

with an axial or flat symmetry, when the boundary  $b(z)$  is a single-valued function,  $|b'(z)|$  is limited, but the boundary variation  $h = \max(b(z) - b)$  is not necessarily small. Substituting the solution — a series in space harmonics with unknown coefficients — into the BC, and expanding it into a complete set gives an (infinite) matrix equation. For  $h \ll b$  it reduces to the  $\varepsilon$ -expansion. The matrix equation is solved numerically after the truncation. The method is convenient for short-period structures like bellows because truncated matrices as small as  $5 \times 5$  are sufficient: the truncation size required is proportional to the period length [29].

**Method of effective boundary conditions** The idea is to impose some *effective BCs*, usually on a smooth surface, which simplify the solution for the fields. One example is “impedance BCs” in a cylindrical waveguide of radius  $b$  with the rectangular corrugation having pitch  $D$  and depth  $h$  ( $h \ll b$ )

$$E_z = iZ_0 \left(1 - \frac{g}{D}\right) \tan\left(\frac{\omega h}{c}\right) H_\varphi; \quad E_\varphi = 0; \quad (2)$$

at  $r = b$ , where  $g/D$  is the filling coefficient [30]. The other effective BCs, obtained by conformal mapping of the corrugation cell into a rectangle [31], replace the corrugated surface by an equivalent smooth one at  $r = b_{eff}$  having a “dielectric coating” with permittivity depending on  $z$ . The high-frequency impedance resonance due to a synchronous wave slowed down by the corrugations was analytically calculated [31] with this approach.

A version of the *effective BCs* approach was applied for calculating the coupling impedances of small pumping holes in the chamber walls [32, 33] for wavelengths larger than a typical hole size. The method, based on the Bethe theory of diffraction by small holes [34], replaces the hole by effective “magnetic” currents to satisfy BCs. For small holes, the fields radiated by these currents are approximated by those due to electric and magnetic dipoles, which allows an analytical evaluation of the impedances. The approach works for all small discontinuities, and the impedance and parameters of the trapped modes (Sec.3.2.7) are expressed in terms of the electric and magnetic polarizabilities of the discontinuity [35]. The polarizabilities have been calculated for holes in a wall of a finite thickness by the variational method [36], for axisymmetric cavities and irises using conformal mapping [37], and for 3-

D obstacles protruding into the chamber like a post or a mask [38]. Another example of using impedance BCs for calculating the impedances of perforated beam pipes is [39].

**Integral-equation methods** The impedance calculation for a discontinuity in an infinite pipe can be reduced to the solution of an integral equation. For an axisymmetric cavity an integral equation was derived [40, 41], in which the unknown function is the field at the cavity-pipe boundary, and the kernel is a series over the cavity EFs. An integral over its solution gives the impedance. When the discontinuity is small, the impedance is expressed in an analytical form [42]. The same is possible for some cavities in the high-frequency limit. The results can also be generalized to several discontinuities and, in the limit, to periodic structures [41].

The impedance can frequently be calculated as a *variational* expression in the unknown field in the integral equation. Using a truncated series with arbitrary coefficients for the unknown field leads to rapidly convergent numerical results for the impedance, e.g., [24, 36].

Another integral equation was obtained for axisymmetric discontinuities of a rather general shape, with the only limitation that  $b(z)$  is a single-valued function [43]. The impedance is expressed as its solution at a particular point. For simple boundaries the kernel can be evaluated analytically, but in the general case it is a definite integral. A more refined development [44] leads to an integro-algebraic equation, which works better above the pipe cut-off.

The impedance calculation for a cylindrical semi-infinite waveguide with perfectly conducting walls was reduced to a solution of a set of two integral equations [45, 46], which can be solved by the factorization technique.

**Diffraction methods** Methods of the *diffraction theory* are used to calculate impedances at high frequencies,  $\omega b/c \gg 1$ , where  $b$  is a typical transverse size of the chamber, following the pioneer work [47]. A contradiction between two results on the high-frequency impedance behavior — either  $\text{Re } Z(\omega) \sim \omega^{-1/2}$  [47, 14] or  $\sim \omega^{-3/2}$  [48] — existed for a long time. It has been resolved first by the matching technique [49]: for a single pill-box cavity the high-frequency behavior  $\text{Re } Z(\omega) \sim \omega^{-1/2}$  holds, whereas for an infinite periodic structure  $\text{Re } Z(\omega) \sim \omega^{-3/2}$ . For a finite number  $M$  of cavities there is a transi-

tion from  $\omega^{-3/2}$ -behavior (which takes place in the frequency range  $L/b \ll \omega b/c \ll M^{2/3}L/b$ ) to  $\omega^{-1/2}$  when  $\omega b/c \gg ML/b$ , where  $L$  is the distance between elements. The result was confirmed by the diffraction model in [50], as well as in [41], where the asymptotic behavior of  $\text{Im } Z$  was also studied, see [1].

**Other methods** A few methods which solve the Maxwell equations in the frequency domain directly do not fit the above classification. They include analytical calculations of the impedance of a smooth chamber at finite beam energy (the space-charge impedance) [51], the resistive-wall impedance for an axisymmetric chamber [52] or the chamber with an arbitrary cross section [53], and the impedance due to the curvature of a smooth chamber [54]. One should mention here two proofs that the impedances of discontinuities asymmetrical in the longitudinal direction are independent of the beam direction [40, 55].

The impedance of an arbitrarily-shaped cavity with semi-infinite pipes can be calculated using known frequency-domain codes for field computations [56], with the radiation BCs simulated by introducing a medium with a small imaginary permittivity, which fills the smooth pipe. Computations, however, are to be performed separately for each frequency.

**Recent developments** The developments have been mostly driven by requirements of projects under construction and in design stage. For low-energy high-intensity accelerators, it is important to know the impedances of non-ultrarelativistic (fixed  $\beta = v/c < 1$ ) beams. The  $\beta$ -dependence was studied analytically for the impedances of small holes [57, 58] and small discontinuities of arbitrary shape [59], for the resistive-wall wakes and impedances [60], and for loss factors [61]. A detailed knowledge of the impedances at very low frequencies is important for the beam stability in large hadron colliders. The calculations of the low-frequency impedances in multi-layer chambers [62, 63, 64, 65], based on the field matching and including skin effects, improve upon the earlier results, see in [7]. High-energy electron linacs and colliders, as well as x-ray FELs, require very short bunches, which stimulated renewed interest in short-range wakes and corresponding very-high-frequency impedances. The high-frequency impedances have been calculated for the wall roughness [66, 67], small wall corrugations [68, 69], and resistive wall [70] including the

anomalous skin effect [71]. A synchronous mode [68, 69] with  $f_{\text{res}} \propto 1/\sqrt{h}$ , where  $h$  is the corrugation (roughness) depth, is similar to one found in a dielectric-layer model [72]; it can be related to an even earlier result [31] obtained using the effective BCs. The high-frequency impedances of a finite-length resistive insert into a smooth pipe were calculated in [73, 74]; various frequency regimes are studied in [75]. The impedances of tapered transitions were originally derived for an axisymmetric case in the low-frequency approximation [76]. More recently, the perturbation corrections were calculated [77], and the derivations were extended to the rectangular [78] and elliptical [79] geometry. Many recent results are reviewed in [8]. Related to the impedance, the problem of beam shielding has been addressed — shielding by an array of concentric wires [80] and by thin conductors [81].

**Codes** Some codes used for impedance calculations are listed in Tab.1; for an extensive list of older and specialized codes see [82]. The table gives the vacuum chamber layout, where the code is applicable, and the code method in frequency ( $f$ -) or in time ( $t$ -) domain;  $W \rightarrow Z$  means finding impedances  $Z$  by Fourier transforming computed wakes  $W$ . Eigensolvers find frequencies and fields of resonant modes, from which  $R_{sh}/Q$  and loss factors can be computed, e.g. [61], usually in post processing. The ‘Output’ column shows the code output: the longitudinal ( $L$ ) or transverse ( $T$ ) impedance  $Z$  or loss factor  $k$ ;  $R/Q$  refers to the impedance of resonant modes and  $Z(f)$  means the impedance in some frequency range. The code references include their web addresses; codes available free are marked by \*. The symbol ‘p’ shows the codes with parallel versions implemented, FIT stands for Finite Integration Technique (Sec.3.2.3), FD for Finite Difference, and FE for Finite Element methods. Time-domain codes are generally more universal and applicable for various discontinuities. They compute wake potentials created by a bunch of a fixed shape. However, one can encounter difficulties in calculating high-frequency impedances due to a limited bunch spectrum: wakes of short bunches are needed to accurately calculate impedances at high frequencies, while computing such wakes requires very small mesh steps. Code NOVO [83] calculates the wakes of very short bunches more efficiently due to its implicit scheme for solving field equations. On the other hand, to resolve the

Table 1: Some codes for impedance computations (see text for symbol definitions).

| Layout                                     | Code       | Method                       | Output                | Reference |
|--|------------|------------------------------|-----------------------|-----------|
| periodic axisymm.<br>rectang. corrugations | KN7C       | $f$ , matching               | $f_{\text{res}}, k_L$ | [20]*     |
|  | TRANSVRS   | $f$ , matching               | $f_{\text{res}}, k_T$ | [21]*     |
|  | Superfish  | $f$ , EF                     | $f_{\text{res}}, R/Q$ | [15]*     |
| axisymmetric<br>discontinuities            | ABCI       | $t, W \rightarrow Z$         | $Z_{L,T}(f), k(f)$    | [86]p*    |
|  | NOVO       | $t, W \rightarrow Z$         | $W_{L,T}(s)$          | [83]*     |
|  | CST MWS(E) | $f$ , EF, FIT                | $f_{\text{res}}, R/Q$ | [16]p     |
|  | CST PS     | $t$ , FIT, $W \rightarrow Z$ | $Z_{L,T}(f)$          | [16]p     |
| general<br>discontinuities                 | GdfidL     | $t$ , FD, $W \rightarrow Z$  | $Z_{L,T}(f)$          | [87]p     |
|  | HFSS       | $f$ , EF, FE                 | $f_{\text{res}}, R/Q$ | [88]p     |

impedances at very low frequencies, one has to compute very long wakes, e.g. [84], which can cause the computations to become numerically unstable. A new computational scheme [85] that eliminates numerical dispersion in the beam direction and therefore is more stable for calculating long wakes was implemented in the code ECHO.

## References

- [1] S. Heifets, S. Kheifets, RMP 63 (1991) 631
- [2] S.S. Kurennoy, Phys. Part. Nucl. 24, (1993) 380; CERN SL/91-31(AP)(1991)
- [3] R.L. Gluckstern, PAC 93, 3219; CERN-2000-011 (2000)
- [4] A.W. Chao, Physics of Collective Beam Instabilities in High Energy Accelerators, Wiley (1993)
- [5] K.-Y. Ng, AIP Proc. 184 (1989) 472; P. Wilson, *ibid* 525
- [6] L. Palumbo, V.G. Vaccaro, M. Zobov, Proc. CAS (1993), CERN 95-06, 331
- [7] B. Zotter, S. Kheifets, Impedances and Wakes in High-Energy Particle Accelerators, World Scientific (1998)
- [8] G.V. Stupakov, PAC 09, 4270
- [9] A. Hofmann, Proc. 11th HEACC (1980) 540
- [10] R.E. Shafer, UM-HE 84-1 (1984) 155
- [11] K.-Y. Ng, PA 24 (1988) 93
- [12] K.-Y. Ng, Physics of the Superconducting Super-collider, Snowmass (1986) 592
- [13] G.V. Stupakov, S.S. Kurennoy, PR E49 (1994) 794
- [14] G. Dome, CERN SPS/85-27 (ARF) (1985)
- [15] K. Halbach, R.F. Holsinger, PA 7 (1976) 213; laacg.lanl.gov
- [16] T. Weiland et al, SLAC-303 (1986) 282; see Sec.3.2.3; CST Studio Suite, www.cst.com
- [17] S.S. Kurennoy, S.V. Purtov, IHEP 88-11 (1988)
- [18] E. Keil, B. Zotter, PA 3 (1972) 11
- [19] H. Henke, PA 25 (1990) 183
- [20] E. Keil, NIM 100 (1972) 419; keil.web.cern.ch/keil/Fortran/
- [21] K. Bane, B. Zotter, Proc. 11th HEACC (1980) 581 www.slac.stanford.edu/grp/ara/groups/BIG/codes.shtml
- [22] L. Vos, CERN SPS/86-21(MS) (1986)
- [23] S. Kheifets, K. Bane, H. Bizek, SLAC-PUB-4097 (1987)
- [24] H. Okamoto, S. Jiang, R.L. Gluckstern, PR E50 (1994) 1501; PA 52 (1996) 179
- [25] S. Heifets, S. Kheifets, SLAC-PUB-3965 (1986)
- [26] S. Krinsky, Proc. 11th HEACC (1980) 576
- [27] S. Krinsky, R. Gluckstern, PAC 81, 2621
- [28] S. Kheifets, B. Zotter, NIM A243 (1986) 13
- [29] S.S. Kurennoy, S.V. Purtov, PA 34 (1990) 155
- [30] B.Z. Katzenelenbaum, High-Frequency Electrodynamics, Nauka (1966)
- [31] V.I. Balbekov, IHEP 85-128 (1985)
- [32] S.S. Kurennoy, PA 39 (1992) 1
- [33] R.L. Gluckstern, PR A46 (1992) 1106; 1110
- [34] H.A. Bethe, PR 66 (1944) 163
- [35] S.S. Kurennoy, R.L. Gluckstern, G.V. Stupakov, PR E52 (1995) 4354
- [36] R.L. Gluckstern J.A. Diamond, IEEE Trans. MTT 39 (1991) 274
- [37] S.S. Kurennoy, G.V. Stupakov, PA 45 (1994) 95
- [38] S.S. Kurennoy, PR E55 (1997) 3529
- [39] S. Petracco, PR E60 (1999) 6030
- [40] R.L. Gluckstern, B. Zotter, CERN LEP-Note 613 (1988)
- [41] R.L. Gluckstern, PR D39 (1989) 2773; 2780
- [42] R.L. Gluckstern, F. Neri, PAC 89, 1271
- [43] S.S. Kurennoy, S.V. Purtov, PA 36 (1992) 223
- [44] R.L. Warnock, SLAC-PUB-6038 (1993)
- [45] S. Heifets, L. Palumbo, V.G. Vaccaro, CERN-LEP-TH/85-23 (1985)
- [46] S. Heifets, L. Palumbo, CERN LEP Note 580 (1987)
- [47] J.D. Lawson, Report RHEL/M144 (1968); PA 25 (1990) 107

### Sec.3.2: IMPEDANCES AND WAKE FUNCTIONS

- [48] K. Bane, P. Wilson, Proc. 11th HEACC (1980) 592
- [49] S. Heifets, S. Kheifets, PR D39 (1989) 960
- [50] S.A. Heifets, PR D40 (1989) 3097
- [51] C.E. Nielsen, A.M. Sessler, K.R. Symon, Proc. Int. Conf. High Energy Acc. (1959) 239
- [52] V.K. Neil, A.M. Sessler, RSI 36 (1965) 429
- [53] R.L. Gluckstern, J.B.J. van Zeijts, B. Zotter, PR E47 (1993) 656
- [54] R.L. Warnock, P. Morton, PA 25 (1990) 113; K.-Y. Ng, *ibid* 153
- [55] S.A. Heifets, SLAC/AP-79 (1990)
- [56] R.L. Gluckstern, F. Neri, Proc. HEACC (1987) v.2, 170
- [57] S. De Santis et al, PR E54 (1996) 800
- [58] A.M. Al-Khateeb et al, J. Phys. G 27 (2001) 2471
- [59] S.S. Kurennoy, PRST-AB 9 (2006) 054201
- [60] F. Zimmermann, K. Oide, PRST-AB 7, 044201 (2004)
- [61] S.S. Kurennoy, PRST-AB 2,014201 (1999)
- [62] A. Burov, V. Lebedev, EPAC 2002, 1452
- [63] E. Mérat, B. Zotter, B. Salvant, PAC 07, 4216
- [64] M. Ivanyan et al, PRST-AB 11, 084001 (2008)
- [65] H. Hahn, PRST-AB 13, 012002 (2010)
- [66] G.V. Stupakov, PRST-AB 1, 064401 (1998)
- [67] G.V. Stupakov et al, PRST-AB 2, 060701 (1999)
- [68] A. Mostacci et al, PRST-AB 5, 044401 (2002)
- [69] K. Bane, G.V. Stupakov, PRST-AB 6, 024401 (2003)
- [70] K. Bane, G.V. Stupakov, PAC 05, 3390
- [71] B. Podobedov, PRST-AB 12, 044401 (2009)
- [72] A.V. Novokhatsky, A. Mosnier, PAC 97 (1997) 1661
- [73] S. Krinsky, B. Podobedov, R.L. Gluckstern, PRST-AB 7, 114401 (2004)
- [74] G.V. Stupakov, PRST-AB 8, 044401 (2005)
- [75] Y. Shobuda, Y.H. Chin, K. Takata, PRST-AB 12, 094401 (2009)
- [76] K. Yokoya, CERN SL/90-88(AP) (1990)
- [77] B. Podobedov, S. Krinsky, PRST-AB 9, 054401 (2006)
- [78] G.V. Stupakov, PRST-AB 10, 094401 (2007)
- [79] B. Podobedov, S. Krinsky, PRST-AB 10, 074402 (2007)
- [80] T.-S.F. Wang, S.S. Kurennoy, R.L. Gluckstern, PRST-AB 4, 104201 (2001)
- [81] R.L. Gluckstern, B. Zotter, PRST-AB 4, 024402 (2001)
- [82] Los Alamos Accelerator Code Group, LA-UR-90-1766 (1990)
- [83] A.V. Novokhatski, SLAC-PUB-11556 (2005); [www.slac.stanford.edu/~novo/](http://www.slac.stanford.edu/~novo/)
- [84] S.S. Kurennoy, EPAC 2002, 1526
- [85] I. Zagorodnov, T. Weiland, PRST-AB 8, 042001 (2005)
- [86] Y.H. Chin, PAC 93, 3414; <http://abci.kek.jp/abci.htm>
- [87] W. Bruns, PAC 97, 2651; [www.gdfidl.de](http://www.gdfidl.de)
- [88] High Frequency Structure Simulator, [www.ansys.com](http://www.ansys.com)

#### 3.2.3 Impedance Calculation. Time Domain E. Gjonaj, T. Weiland, TU Darmstadt

The primary quantity of interest in time domain calculations is the longitudinal wake potential,

$$W_{||}(x, y, s) = -\frac{1}{q} \int_{-\infty}^{\infty} dz E_z(x, y, z, \frac{s+z}{c}), \quad (1)$$

where  $q$  is the charge of a wakefield inducing bunch and  $s$  is the relative position of a witness particle with respect to the bunch. Coupling impedances and other related parameters, such as transverse kick factors are determined from  $W_{||}$  (e.g., using the Panofsky-Wenzel theorem [1]).

The calculation of (1) is based on the transient solution of the full set of Maxwell's equations. Usually, no simplifying assumptions are involved which would limit the applicability of the method. Accelerator structures of arbitrary 3D-geometry with resistive and rough surface walls as well as bunches of arbitrary distribution can be handled. The drawback resulting from such a general approach is the large numerical effort which is often needed in terms of simulation time and memory needs. In the important case of the short-range coupling of ultra-relativistic bunches, however, specialized numerical algorithms exist which make time domain calculations feasible for most practical purposes. In the following, a brief description of these algorithms is given.

**Space and time discretizations** The starting point is a set of semidiscrete (time-continuous-space-discrete) matrix equations of the form,

$$\frac{d}{dt} \begin{pmatrix} M_e e \\ M_\mu h \end{pmatrix} = \begin{pmatrix} 0 & C^T \\ -C & 0 \end{pmatrix} \cdot \begin{pmatrix} e \\ h \end{pmatrix} - \begin{pmatrix} j \\ 0 \end{pmatrix}, \quad (2)$$

where  $e$ ,  $h$  denote discrete electromagnetic degrees of freedom and  $j$  is the discrete representation of the bunch current on a computational grid. For the sake of simplicity, conducting currents induced in accelerator structures of finite electrical conductivity are neglected.

The matrices  $M_\epsilon$ ,  $M_\mu$  and  $C$  in (2) represent the discrete *mass*- and *curl*-operators, respectively, as obtained by the application of an appropriate discretization method. The traditional method used for this purpose is the Finite Integration Technique (FIT) on Cartesian grid [2]. The method provides a second order accurate and numerically highly efficient framework for the solution of Maxwell's equations in the time domain. A number of major 3D-wakefield codes, including MAFIA [3], CST PARTICLE STUDIO® (CST PS) [3], GdfidL [4], ECHO [5] and PBCI [6] are based on this formulation. Furthermore, wakefield calculations using Finite Element-type Methods (FEM) of higher order accuracy on unstructured grids have been reported. This discretization approach is implemented in the code Tau3P [7] developed at SLAC.

Equations (2) are further discretized in time using an explicit (or implicit) time update scheme. The *explicit* update can be generally written as

$$\begin{pmatrix} e \\ h \end{pmatrix}^{n+1} = \mathbf{G}(\Delta t) \cdot \begin{pmatrix} e \\ h \end{pmatrix}^n - \begin{pmatrix} \Delta t M_\epsilon^{-1} j \\ 0 \end{pmatrix}^n, \quad (3)$$

where  $n$  and  $n + 1$  are two subsequent time levels in the computation and  $\mathbf{G}$  is a scheme-specific time update operator in  $(t^n, t^{n+1}]$  depending on the discrete operators  $M_\epsilon$ ,  $M_\mu$  and  $C$ . A numerically efficient and, thus, commonly used update scheme is the leap-frog integrator [8]. This approach is used in the wakefield codes MAFIA and CST PS.

It has been early recognized that, when the leap-frog integrator is applied, the maximum numerical dispersion error occurs for plane wave solutions propagating parallel to the grid axes (cf. [9]). This property represents a drawback of the scheme, since the dominant wakefield propagation direction coincides with the main grid axis,  $z$ , which is commonly aligned with the bunch direction of motion. The resulting numerical error is comparatively large, especially, in calculations involving short bunches. Furthermore, the scheme imposes a stability condition which restricts the time step below the longitudinal periodicity implied by the grid,  $\Delta t < \Delta t_z = \Delta z/c$ . A moving window implementation (see below) based on a leap-frog update is, thus, impracticable. The situation improves when so called *dispersion free schemes* are applied. These are described in the following.

**Dispersion free schemes** The purpose of these schemes is to suppress numerical dispersion for plane wave solutions propagating in the longitudinal direction. Two classes of dispersion free schemes have been reported. The first includes methods which modify the time update (3), while leaving the discrete operators in (2) unchanged. In [9], this is realized by introducing an approximate splitting of the update operator  $\mathbf{G}$  into longitudinal and transversal updates. The Longitudinal-Transversal (LT) splitting reads

$$\mathbf{G}(\Delta t) \approx \mathbf{G}_T \left( \frac{\Delta t}{2} \right) \cdot \mathbf{G}_L(\Delta t) \cdot \mathbf{G}_T \left( \frac{\Delta t}{2} \right), \quad (4)$$

where  $\mathbf{G}_T$  and  $\mathbf{G}_L$  contain only spatial derivatives in the transversal and in the longitudinal directions, respectively. The LT-scheme is shown to be stable and longitudinally dispersion free at the “magic time step”,  $\Delta t = \Delta z/c$ . This is demonstrated in Fig.1, where numerical phase velocities in the  $zx$ -plane for different propagation angles are shown. The LT-scheme is implemented in the 3D-wakefield codes PBCI and GdfidL.

Another approach based on operator splitting is the TE/TM scheme [10, 11]. The explicit formulation introduced in [11] uses the splitting,

$$\mathbf{G}(\Delta t) \approx \mathbf{G}_{TE}(\Delta t) \cdot \mathbf{G}_{TM}(\Delta t), \quad (5)$$

where  $\mathbf{G}_{TE}$  and  $\mathbf{G}_{TM}$  are (partial) time update operators affecting only TE- and TM-wave solutions, respectively. Also this scheme is stable and longitudinally dispersion free at the “magic time step”. When a uniform Cartesian grid is used, the method is dispersion free along the two transversal directions as well. The TE/TM scheme is implemented for 2D- and 3D-wakefield calculations in the code ECHO. In Fig.2, longitudinal wake potentials in a TESLA cavity computed with

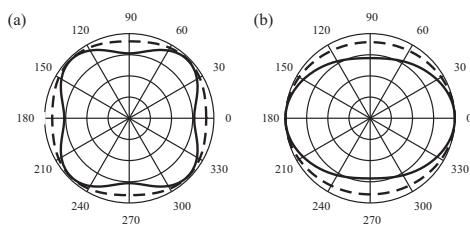


Figure 1: Normalized phase velocities in the  $zx$ -plane at the stability limit for the (a) leap-frog scheme and (b) LT-splitting. Solid and dashed lines represent different grid resolutions.

### Sec.3.2: IMPEDANCES AND WAKE FUNCTIONS

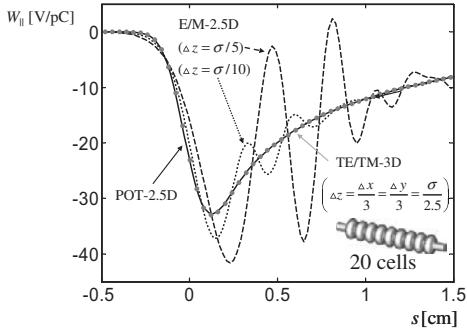


Figure 2: Longitudinal wake potentials presented in [10] for an electron bunch with  $\sigma_z = 1$  mm in a 20-cell TESLA structure. The solutions obtained for different grid resolutions using the leap-frog scheme (dashed lines) and the TE/TM splitting (solid lines), respectively, are shown.

ECHO are shown. The immense advantage of the scheme with respect to numerical accuracy compared to the classical leap-frog approach is clearly observed.

In a second class of dispersion free methods the spatial discretization represented by  $M_\epsilon$ ,  $M_\mu$  and  $C$  is constructed such that the numerical dispersion is minimum in the longitudinal direction, even when a leap-frog update is applied. A Finite Volumes (FV) based discretization method using this idea is reported in [12]. The method is stable at the “magic time step” as well as dispersion free along the three grid axes. The non-split, dispersion free FV-method is implemented in the code PBCI.

**Moving window** When interested in the head-tail interaction of ultra-relativistic bunches, causality implies that the integration range in (1) can be restricted within a small comoving frame including the particle bunch. This idea was first expressed in [13] and it has since then been used in time domain calculations to reduce computational effort. The actual computational domain consists of a small window which is shifted at the speed of light in vacuum to account for the bunch motion.

In order for this approach to work, however, the longitudinal numerical phase velocity must exactly match the speed of light in vacuum. Otherwise, numerical reflections at the window boundaries occur which sooner or later pollute the wakefield solution in the vicinity of the bunch.

This behavior is observed even when approximate absorbing boundary conditions are applied. Dispersion free schemes at the “magic time step” are able to cope with the problem in a straightforward manner. No special boundary conditions are required for terminating the window. Grid and field solution can be simply shifted by  $\Delta z$  in every time step without loss of accuracy. Moving window implementations based on dispersion free schemes are realized in the 3D-codes ECHO, PBCI and GdfidL.

**Boundary conformal methods** Impedance calculations in the time domain are mostly based on a Cartesian grid discretization. This approach is numerically efficient and it enables the application of dispersion free schemes. However, the numerical error resulting from the crude approximation of curved cavity boundaries (staircasing error) may become large. The idea of boundary conformal methods is to improve calculation accuracy by including additional information on the boundary geometry in the spatial discretization step.

A well established boundary conformal method in the general context of electromagnetic field computation is the Partially Filled Cell (PFC) approximation by Thoma-Weiland [14]. The method is implemented for impedance calculations in the fixed window code CST PS, and in the moving window code ECHO. The high accuracy of the method compared to the standard staircase approximation is illustrated in Fig.3, which shows wake potentials for a rotationally symmetric collimator structure computed with ECHO-2D.

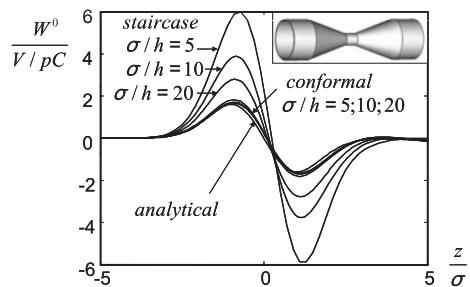


Figure 3: Longitudinal wake potentials for a collimator structure calculated with ECHO-2D. The numerical convergence with respect to grid resolution in staircase and conformal approximation, respectively, is shown.

While the application of a boundary conformal method clearly improves accuracy, it also increases the computational load. This is particularly the case for long accelerator structures, where the operation count needed for extracting detailed geometrical information may become prohibitively large. Another drawback of the method is the time step reduction resulting from stability constraints. This makes the application of dispersion free schemes in this combination difficult. A solution to this problem was found in [15]. The Uniformly Stable Conformal (USC) method implemented in ECHO allows for a conformal discretization without time step reduction.

**Resistive wall wakefields** The numerical techniques mentioned above refer primarily to the wakefields due to accelerator chamber geometry. In certain situations, however, the wakefields induced by the finite conductivity of the chamber walls become important. The effect of these so called resistive wakefields tends to be dominant, in particular, in the “warm” accelerator sections of coherent light sources such as undulators (cf. [16]). In order to handle resistive wakefields numerically, a broadband frequency dependent impedance model for the electrical conductivity of lossy accelerator walls is necessary, which makes the numerical implementation of dispersion free schemes, moving window and boundary conformal approaches extremely cumbersome. The calculation of resistive wakefields is still possible in the fixed-window codes such as MAFIA and CST PS, although the application of these tools is limited to comparatively long bunches and short accelerator structures. Moving window and dispersive free implementations for resistive wakefields remain a challenge for future developments. A recent work presenting such an implementation for rotationally symmetric problems can be found in [17].

**Indirect integration** refers to the procedure used for calculating the wakefield contribution in the potential (1) *after* the bunch has passed the accelerator structure. The distance needed for the wakefields to “catch-up” with the bunch in an outgoing beam pipe behind the structure scales as  $a^2/\sigma$ , where  $a$  is the pipe radius and  $\sigma$  the bunch length. This distance is, usually, several meters long which makes the computation of (1) based on one of the previously described numerical methods impossible. In order to deal with this problem, the wake integral (1) is split into two parts. In the

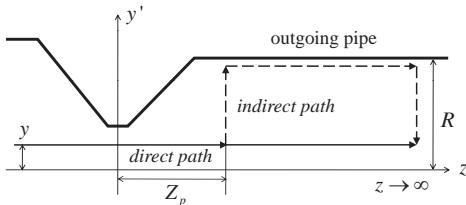


Figure 4: Schematic of the indirect integration path.

range  $(-\infty, Z_p]$ , containing the accelerator structure a (direct) numerical calculation is used. In the outgoing pipe, for  $(Z_p, \infty)$ , a semianalytical procedure (indirect integration) is applied.

A well known indirect integration approach consists in deforming the integration path such that it may be continued to infinity along the conducting wall of the pipe. The situation is illustrated in Fig.4. Using the notations therein, the following relation holds in the general 3D-case:

$$\int_{Z_p}^{\infty} dz E_z(x, y, z, \frac{s+z}{c}) = \int_y^R dy' (E_y - cB_x)^{TM} \left( x, y', Z_p, \frac{s+Z_p}{c} \right). \quad (6)$$

Thus, the improper wake integral can be transformed into a proper one along a short path on the cross-section of the pipe. The relation (6) is due to Henke and Bruns [18]. It represents the generalization of the indirect integration method for rotationally symmetric cavity-like structures which was earlier introduced by Weiland [19] and then further developed by Napoly, Chin and Zotter in [20]. The indirect integration by path deformation is implemented in the 2D-codes BCI, ABCI and TBCI. The 3D-variant given by (6) is implemented in the codes GdfidL and ECHO. Note the subscript  $TM$  appearing in (6) which indicates that, in the 3D-case, the  $TM$ -wave component needs to be extracted from the full field solution in the beam pipe before the indirect integration is performed.

A second approach is based on the decomposition of outgoing wakefields into waveguide modes. Given such a decomposition, the wake integral in  $(Z_p, \infty)$  can be performed analytically. While this approach does not represent an indirect integration in the literal sense, it provides a detailed information on the outgoing wakefields which can be used for purposes other than indirect

### Sec.3.2: IMPEDANCES AND WAKE FUNCTIONS

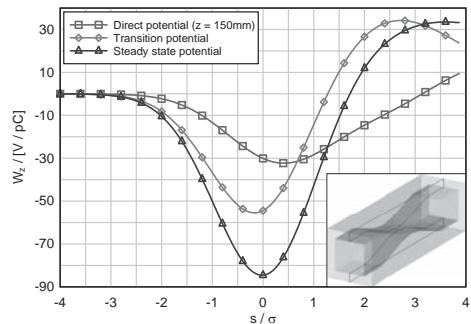


Figure 5: Direct vs. indirect integration for a 0.3 mm-bunch in the ILC collimator #8 [21]. “Transition potential” refers to the wakefield integral contribution in the outgoing pipe. 150 waveguide modes were used in the calculation.

integration (cf. [6]). The indirect integration by modal decomposition is implemented in the code PBCI. Figure 5 illustrates the application of the method in the calculation of the wake potential for one of the rectangular collimators used in the ILC-ESA test beam program [21].

### References

- [1] W.K.H. Panofsky, W.A. Wenzel, RSI 27 (1956) 967
- [2] T. Weiland, AEÜ 31 (1977) 116
- [3] CST AG, <http://www.cst.com>.
- [4] W. Bruns, <http://www.gdfidl.de>
- [5] I. Zagorodnov et al, J.Comp.Phys. 191 (2003) 525
- [6] E. Gjonaj et al, ICAP 06 (2006) 29
- [7] M. Wolf et al, ACES 02 (2002)
- [8] K.S. Yee, IEEE Trans. Antennas and Propagation AP-14 (1966) 302
- [9] T. Lau et al., FREQUENZ 59 (2005) 210
- [10] I. Zagorodnov, T. Weiland, J.Comp.Phys. 207 (2005) 69
- [11] M. Dohlus, I. Zagorodnov, J.Comp.Phys. 228 (2009) 2822
- [12] E. Gjonaj et al., PAC 09 (2009) FR5PFP091
- [13] K. Bane, T. Weiland, HEACC 83 (1983) 314
- [14] P. Thoma, Ph.D. dissertation D17 (1997) TH Darmstadt
- [15] I. Zagorodnov et al, Int.J.Num.Mod. 16 (2003) 127
- [16] K.L.F. Bane, G. Stupakov, PAC 05, 3390
- [17] A. Tsakanian, M. Dohlus, I. Zagorodnov, PAC 09, 3342
- [18] H. Henke, W. Bruns, EPAC 06, 2170
- [19] T. Weiland, NIM 216 (1983) 31
- [20] O. Napoly et al, NIM A334 (1993) 255
- [21] N. Watson et al, EPAC 06, 697

### 3.2.4 Explicit Expressions of Impedances and Wake Functions

K.Y. Ng, FNAL, K. Bane, SLAC

See tables in the next pages.

### References

- [1] A.W. Chao, Physics of Collective Beam Instabilities in High Energy Accelerators, Wiley (1993) ch.2
- [2] K.Y. Ng, Fermilab-FN-0756, 2004
- [3] K. Bane, M. Sands, Proc. Micro Bunches Workshop 1995, 131
- [4] S. Krinsky et al, PRST-AB 7, 114401 (2004); Proc. Micro Bunches Workshop 1995, 131
- [5] A. Piwiniski, DESY 94-068 (1994); DESY 84-097 (1984)
- [6] A. Piwiniski, PAC 77, 1364; X.E. Lin, SLAC-PUB-7924 (1998)
- [7] L. Palumbo and V.G. Vaccaro, Nuovo Cimento A 89 (1985) 3
- [8] R. Gluckstern et al, PR E47 (1993) 656
- [9] K. Yokoya, PA 41 (1993) 221
- [10] A. Lutman et al, PRST-AB 11, 074401 (2008)
- [11] K.Y. Ng, PA 16 (1984) 63
- [12] N. Mounet, E. Métral, CERN-BE-2009-039 (2009) For a Mathematica code to compute multi-layer impedances, visit [https://impedance.web.cern.ch/impedance/Codes/ReWall/ReWall\\_to\\_date.zip](https://impedance.web.cern.ch/impedance/Codes/ReWall/ReWall_to_date.zip) and <https://impedance.web.cern.ch/impedance>
- Note that we separate impedances into the usual space-charge (self-field plus wall image, both  $\rightarrow 0$  as  $\gamma \rightarrow \infty$ ) and wall impedances, whereas Mounet separates them instead into self-direct (due to beam only but excluding wall image) and wall (our usual wall impedance plus wall image).
- [13] B. Zotter, CERN-AB-2005-043 (2005)
- [14] A. Burov, V. Lebedev, EPAC 02 (2002) 1402
- [15] G. Stupakov et al, PRST-AB 10, 054401 (2007); K. Bane et al, PRST-AB 10, 074401 (2007)
- [16] S.A. Heifets, PR D40 (1989) 3097; S.A. Heifets, S.A. Kheifets, Rev. Mod. Phys. 63 (1990) 631
- [17] E. Gianfelice, L. Palumbo, IEEE Tr. NS 37, 2 (1990) 1081
- [18] I. Zagorodnov, K. Bane, EPAC 06, 2859

**General Remarks and Notation:**

In cylindrically symmetric structures  $W'_m(z)$  and  $W_m(z)$  denote, respectively,  $m$ -th azimuthal multipole longitudinal and transverse wake functions, generated by point charge  $Q$ , at distance  $-z > 0$  behind.  $W'_m(z) \equiv dW_m(z)/dz$ .  $W'_m(z) = 0$  and  $W_m(z) = 0$  when  $z > 0$  when particle travels at the speed of light.  $W'_m(0) = \frac{1}{2} \lim_{z \rightarrow 0} W'_m(z)$ . Longitudinal and transverse momentum kicks on test charge  $q$  near pipe axis:  $\Delta p_{\parallel}(z) \stackrel{z \rightarrow 0}{=} -qQW'_0(z)/c$ ,  $\Delta p_{\perp}(z) = -qQ\tilde{r}_{\perp}W_1(z)/c$ , where  $\tilde{r}_{\perp}$  is (small) offset of the source or exciting charge.

The  $m$ -th multipole longitudinal impedance  $Z_m^{\parallel}(k) = \int e^{-ikz/\beta} W_m^{\parallel}(z) dz / (\beta c)$  is related to the  $m$ -th multipole transverse impedance,  $Z_m^{\perp}(k) = i \int^* e^{-ikz/\beta} W_m^{\perp}(z) dz / (\beta^2 c)$ , by  $Z_m^{\parallel} = k Z_m^{\perp}$  ( $m \neq 0$ ), where  $k = \omega/c$ . Note that  $Z_m^{\parallel}(-k) = Z_m^{\parallel}(k)$ ,  $Z_m^{\perp}(-k) = -Z_m^{\perp*}(k)$ .

For periodic or translationally invariant structures: steady-state results are given per length  $L$ . Unless otherwise stated, structures are cylindrically symmetric with perfectly conducting metallic walls, and with beam pipes of radius  $b$ . In many cases,  $\beta = v/c$  has been set to 1.  $Z_0 = \sqrt{\mu_0/\epsilon_0} \approx 377\Omega$  is impedance,  $\epsilon_0$  electric permittivity, and  $\mu_0$  magnetic permeability of free space. ‘Pill-box cavity’ signifies a pill-box with beam pipes. Here  $[\alpha \pm i|\beta|]^{1/n}$  (with  $\alpha, \beta$  real,  $n=2$  or 3) is in the 1st/4th quadrant.  $H(x) = 0, 1$  for  $x \leq 0$ .

For 3D structures with mirror symmetry in  $x$  and  $y$ , near axis momentum kick in  $y$ ,  $\Delta p_y = -qQ(\tilde{y}W_d^y + yW_q^y)$ , with  $\tilde{y}$  ( $y$ ) offset of exciting (test) charge, and  $W_d^y$  ( $W_q^y$ ) dipole (quad) wake terms. Total  $y$  wake  $W_y = W_d^y + W_q^y$ ; total  $y$  impedance  $Z_y = Z_d^y + Z_q^y$ .

| Description   | Impedances  |  | wake   |   |
|---|---|--|--|---|
| <b>Space-charge:</b> [1]<br>beam radius $a$ in a perfectly conducting beam pipe of radius $b$ , transverse distribution uniform.  | $\frac{Z_0^{\parallel}}{L} = i \frac{Z_0 k g_0}{4\pi \beta^2 \gamma^2}$ $g_0 = 1 + 2 \ln \frac{b}{a}$   | $\frac{W'_0}{L} = \frac{Z_0 c}{4\pi \gamma^2} \left[ 1 + 2 \ln \frac{b}{a} \right] \delta'(z)$ | $\frac{Z_0^{\perp}}{L} = i \frac{Z_0}{2\pi \beta^2 \gamma^2 m} \left[ \frac{1}{a^{2m}} - \frac{1}{b^{2m}} \right]$ | $\frac{W_{m \neq 0}}{L} = \frac{Z_0 c}{2\pi \gamma^2 m} \left[ \frac{1}{a^{2m}} - \frac{1}{b^{2m}} \right] \delta(z)$ |
| Nonuniform distributions: [2] $a_{\text{eff}}^{\parallel}$ is equivalent-uniform-beam radius, $g_0 = 1 + 2 \ln(b/a_{\text{eff}}^{\parallel})$ , while $a_{\text{eff}}^{\perp} = [\pi \lambda(0)]^{-1/2}$ is the same when self-force part written as $1/a_{\text{eff}}^{\perp 2}$ , $\gamma_e \approx 0.57721$ is Euler's constant. | Distribution $\lambda(r)$   | $g_0$ ( $m=0$ )  | $a_{\text{eff}}^{\parallel}$ ( $m=0$ )   | $a_{\text{eff}}^{\perp}$ ( $m=1$ )  |
|   | $\frac{3}{2\pi \hat{r}} \left( 1 - \frac{r^2}{\hat{r}^2} \right)^{1/2} H(\hat{r}-r)$  | $\frac{8}{3} + 2 \ln \frac{b}{2\hat{r}}$   | $0.8692\hat{r}$  | $\sqrt{\frac{2}{3}}\hat{r}$   |
|   | $\frac{1}{2\pi \hat{r}^2} \left( 1 - \frac{r^2}{\hat{r}^2} \right) H(\hat{r}-r)$  | $\frac{3}{2} + 2 \ln \frac{b}{\hat{r}}$  | $0.7788\hat{r}$  | $\frac{1}{\sqrt{2}}\hat{r}$   |
|   | $\frac{2\pi}{\pi^2 - 4} \cos^2 \frac{\pi r}{2\hat{r}} H(\hat{r}-r)$   | $1.921 + 2 \ln \frac{b}{\hat{r}}$  | $0.6309\hat{r}$  | $\frac{\sqrt{\pi^2 - 4}}{\sqrt{2}\pi} \hat{r}$  |
|   | $\frac{1}{2\pi \sigma_r^2} e^{-r^2/(2\sigma_r^2)}$  | $\gamma_e + 2 \ln \frac{b}{\sqrt{2}\sigma_r}$  | $1.7647\sigma_r$   | $\sqrt{2}\sigma_r$  |
| Image part of $Z_1^{\perp}$ can be written in terms of Laslett's electric image coefficients as $1/b^2 \rightarrow 2(\xi_{1x,y} - \epsilon_{1x,y})/h^2$ with $h$ denoting half height of vacuum chamber. See Sec.2.4.5.   |   |  |  |   |
| <b>Resistive wall:</b> [1, 3]<br>wall thickness $t$ , dc and ac conductivities $\sigma_c$ , $\tilde{\sigma}_c$ , relaxation time $\tau$ ; assume $ k b \gg (s_0/b)^3$ , thick walls: $t \gg \delta_c = \sqrt{2}/( k Z_0\mu_r\sigma_c)$ , the skin depth.  | $\frac{Z_m^{\parallel}}{L} = \frac{Z_0/(pb^{2m+1})}{(1+\delta_m0)\sqrt{\frac{iZ_0\tilde{\sigma}_c}{k\mu_r}} - \frac{ibk}{m+1}}$   |  | $\mu_r = \begin{cases} \text{relative magnetic} \\ \text{permeability} \end{cases}$                                | $\tilde{\sigma}_c = \sigma_c/(1-ik\tau)$  |
|   | Typically, $\tau = 27/40/8$ fs for Cu/Ag/Al. Valid for $c\tau/s_0 \ll 1$ , characteristic distance $s_0 = [2b^2\mu_r/(Z_0\sigma_c)]^{1/3}$ , $\alpha = [(m+1)(1+\delta_m0)/2]^{2/3}$ .  |  |  |   |
|   | $\frac{W'_m}{L} = \frac{4Z_0 c (m+1)}{\pi b^{2m+2}} \left[ \frac{e^{\alpha z/s_0}}{3} \cos \left( \frac{\sqrt{3}\alpha z}{s_0} \right) - \frac{\sqrt{2}}{\pi} \int_0^\infty dx \frac{x^2 e^{\alpha zx^2/s_0}}{x^6 + 8} \right]$ |  |  |   |

| Description   | Impedances  |  | Wakes                              |  |  |
|---|---|--|------------------------------------|--|--|
| Low frequency: [1]<br>$k \ll 1/s_0$ , long range<br>$ z  \gg s_0$ .   | $Z_m^{\parallel} = \frac{1 - \text{sgn}(k)i}{(1 + \delta_{m0})\pi\sigma_c\delta_c b^{2m+1}}$<br>Note : $Z_1^{\perp} = \frac{2}{b^2 k} Z_0^{\parallel}$ ,  | $\frac{W'_m}{L} = \frac{-c}{2\pi b^{2m+1}(1 + \delta_{m0})} \sqrt{\frac{Z_0\mu_r}{\pi\sigma_c}} \frac{1}{(-z)^{3/2}}$  | $W_1 = \frac{2}{b^2} \int W'_0 dz$ |  |  |
| Low frequency, thin wall: [1] $t \ll \delta_c$ and $ k  \ll 1/\sqrt{bt}$ .  | $Z_0^{\parallel} = -\frac{Z_0 kt}{2\pi b}, \quad Z_1^{\perp} = -i \frac{Z_0 t}{\pi b^3}$  | $\frac{W'_0}{L} = -\frac{Z_0 tc}{2\pi b} \delta'(z), \quad \frac{W_1}{L} = -\frac{Z_0 ct}{\pi b^3} \delta(z)$  |                                    |  |  |
| High frequency: [3]<br>$k \gtrsim 1/s_0$ , short range<br>$ z  \lesssim s_0$ , with $c\tau \gtrsim s_0$ .<br>$k_p = \sqrt{Z_0\sigma_c/c\tau}$ is plasma frequency/ $c$ .  | $Z_m^{\parallel} = \frac{4Z_0 c\tau(m+1)}{\pi b^{2m+1}}$<br>$\times \frac{1 - 4ikc\tau}{b(1 - 4ikc\tau)^2 + 32k_p(\alpha c\tau)^2}$   | $\frac{W'_m}{L} = \frac{Z_0 c(m+1)}{\pi b^{2m+2}} e^{\frac{iz}{4c\tau}}$<br>$\times \cos \left[ \sqrt{\frac{2k_p}{b}} \alpha z \right]$ , for $\alpha$ see above |                                    |  |  |
| Finite length, lossy insert: [4] of length $L$ , in lossless pipe   | These formulae depend only on the plasma frequency of the metal. Effects of relative magnetic permeability have not been considered.  |  |                                    |  |  |
| Displaced beam: [5] at $\vec{a} = (a_x, a_y)$ , rms bunch length $\sigma_\ell$ , average current $I_b$ , and $(b/k^2, b, b-a) \gg \delta_c$ and $\gamma \gg 1$ .  | Wall impedances in last section multiplied by $f_z$ for $Z_0^{\parallel}$ and $f_{x,y}$ for $Z_1^{\perp}$ with $f_z = \frac{b^2 + a^2}{b^2 - a^2}$ , $f_x = \frac{b(b^2 - a^2 + 4a_x^2)}{(b^2 - a^2)^3}$ , $f_y = \frac{b(b^2 - a^2 + 4a_y^2)}{(b^2 - a^2)^3}$ Power loss per length traversed is $\frac{P}{L} = \frac{\Gamma(\frac{3}{4})I_b^2}{4\pi^2 b\sigma_\ell^{3/2} \sqrt{2\mu_r\sigma_c/Z_0}} f_z$  |  |                                    |  |  |
| Displaced beam between two infinite plates: [5] at $y = \pm h/2$ , $\gamma \gg 1$ , $[h/k^2, h-2y_0] \gg \delta_c$ .  | $Z_0^{\parallel} = \frac{1 - \text{sgn}(\omega)i}{\pi h} \sqrt{\frac{ \omega \mu_r Z_0}{2c\sigma_c}} f_z, \quad Z_1^{\perp} = \frac{\pi(\text{sgn}(\omega)1-i)}{\sqrt{2 \omega \sigma_c/(c\mu_r Z_0)}} f_{\perp}$<br>$f_z = 1 + \frac{\pi y_0}{h} \tan \frac{\pi y_0}{h}, \quad f_{\perp} = \frac{f_z}{h^3 \cos^2(\pi y_0/h)}$  | beam at $y = y_0$  |                                    |  |  |
| Thin dielectric coating of thickness $\Delta h$ .   | $Z_0^{\parallel} = -\frac{i\omega Z_0(\epsilon_r\mu_r-1)\Delta h}{\pi c\epsilon_r h} f_z, \quad Z_1^{\perp} = -\frac{i\pi Z_0(\epsilon_r\mu_r-1)\Delta h}{\epsilon_r} f_{\perp}$  |  |                                    |  |  |
| Metallic coating on ceramic pipe: [6] compared with all metal pipe $Z_0^{\parallel}(\text{met})$ . $t_{m,c}$ = metal/ceramic thickness $\ll b$ . $\gamma \gg 1$ , $[(\epsilon_r-1)t_c^2, (1-\epsilon_r^{-1})bt_c] \ll \sigma_\ell^2$ . Loss $P/L$ is max. at $V = 0.82$ . | $Z_0^{\parallel} = Z_0^{\parallel}(\text{met}) \frac{A + \tanh(\nu t_m)}{1 + A \tanh(\nu t_m)}, \quad A = \left(1 - \frac{1}{\epsilon_r}\right) \nu t_c, \quad \nu = \frac{1 - \text{sgn}(\omega)i}{\delta_c}$<br>$\frac{P}{L} = \frac{Z_0 I_b^2 t_c (\epsilon_r - 1)}{4\sqrt{\pi} b \sigma_\ell^2 \epsilon_r} \left[ V - \sqrt{\pi} V^2 e^{V^2} \text{erfc}(V) \right], \quad V = \frac{\epsilon_r \sigma_\ell}{(\epsilon_r - 1) Z_0 \sigma_c t_m t_c}$<br>Field penetration through pipe, $\frac{E_{z,\text{out}}}{E_{z,\text{in}}} = \frac{1}{\sqrt{1 + 4(1 - 1/\epsilon_r)t_m t_c / \delta_c^2}}$ , becomes significant when $t_m \lesssim t_{\text{crit}} = \delta_c^2/t_c$ . $P/L$ is at maximum at $t_{\text{crit}}$ . |  |                                    |  |  |
| Elliptical beam pipe:   | Low frequency, see [7, 8, 5], high frequency, see [9, 10].  |  |                                    |  |  |
| Rectangular beam pipe:  | Low frequency, see [8], high frequency, see [9, 11].  |  |                                    |  |  |

**Multi-layer pipe wall impedances:** [12, 13] Cylindrical beam pipe with  $N$  layers,  $p$ th layer between  $b^{(p-1)} < r < b^{(p)}$  and  $b^{(N)} \rightarrow \infty$ . Layer 1 is vacuum,  $a < r < b^{(1)}$ , with particle beam of charge  $Q$  at  $r=a$  and  $\theta=0$ .  $r < a$  is called the 0-th layer. Each layer has its own wavenumber  $\nu = k\sqrt{1-\beta^2\varepsilon_1\mu_1}$ ,  $k = \omega/v$  and own properties  $\epsilon = \epsilon_0\varepsilon_1 = \epsilon_0\epsilon_r(1+i\tan\vartheta_E) - \frac{\sigma_{dc}}{i\omega(1-i\omega\tau)}$ ,  $\mu = \mu_0\mu_1 = \mu_0\mu_r(1+i\tan\vartheta_M)$ ;  $\vartheta_E, \vartheta_M$  are loss angles,  $\sigma_{dc}$  dc conductivity, and  $\tau$  relaxation time. Actually any frequency dependent  $\epsilon, \mu$ , and conductivity can be assumed. Inside vacuum,  $\nu = k/\gamma$ ; inside conducting metal of skin depth  $\delta_c$ ,  $\nu \approx (1-i)/\delta_c$ . A user-friendly Mathematica code for computation is available [12]. The derivation is outlined briefly below. In terms of Bessel and Kelvin functions,  $m$ th multipole longitudinal fields inside  $p$ -th layer:

$$E_s^{(p)} = \cos m\theta e^{iks} \left[ C_{Ie}^{(p)} I_m(\nu^{(p)}r) + C_{Ke}^{(p)} K_m(\nu^{(p)}r) \right], \quad \vec{E} \text{ is electric field}$$

$$G_s^{(p)} = \sin m\theta e^{iks} \left[ C_{Ig}^{(p)} I_m(\nu^{(p)}r) + C_{Kg}^{(p)} K_m(\nu^{(p)}r) \right], \quad \vec{G} = Z_0 \vec{H}, \quad \vec{H} \text{ is magnetic field}$$

Matching  $E_s, E_\theta, G_s$ , and  $G_\theta$  at boundary  $r=b^{(p)}$  between  $p$ -th and  $(p+1)$ -th layers gives

$$\begin{bmatrix} C_{Ie}^{(p+1)} \\ C_{Ke}^{(p+1)} \\ C_{Ig}^{(p+1)} \\ C_{Kg}^{(p+1)} \end{bmatrix} = M_p^{p+1} \begin{bmatrix} C_{Ie}^{(p)} \\ C_{Ke}^{(p)} \\ C_{Ig}^{(p)} \\ C_{Kg}^{(p)} \end{bmatrix} \xrightarrow{\text{iteratively}} \begin{bmatrix} C_{Ie}^{(N)} \\ C_{Ke}^{(N)} \\ C_{Ig}^{(N)} \\ C_{Kg}^{(N)} \end{bmatrix} = \mathcal{M} \begin{bmatrix} C_{Ie}^{(1)} \\ C_{Ke}^{(1)} \\ C_{Ig}^{(1)} \\ C_{Kg}^{(1)} \end{bmatrix} \quad \text{where } \begin{cases} \mathcal{M} \equiv M_{N-1}^N M_{N-2}^{N-1} \cdots M_1^2 \\ \text{See [12] for explicit expression of } M_p^{p+1} \end{cases}$$

Since the last layer goes to infinity,  $C_{Ie}^{(N)} = C_{Ig}^{(N)} = 0$ . From the beam region,  $C_{Kg}^{(1)} = 0$  and  $C_{Ke}^{(1)} = -ikQZ_0 I_m(ka/\gamma)/[\pi\beta\gamma^2(1+\delta_{m0})]$ , one can easily solve for

$$C_{Ie}^{(1)} \equiv -\alpha_1 \frac{K_m^{(1)}}{I_m^{(1)}} C_{Ke}^{(1)} = -C_{Ke}^{(1)} \frac{\mathcal{M}_{12}\mathcal{M}_{33} - \mathcal{M}_{32}\mathcal{M}_{13}}{\mathcal{M}_{11}\mathcal{M}_{33} - \mathcal{M}_{13}\mathcal{M}_{31}}, \quad \text{with } I_m^{(1)} = I_m(\nu^{(1)}b^{(1)}), \quad I_m^{(1)} = I_m(\nu^{(1)}b^{(1)}).$$

With beam at  $r=a_1, \theta=0$ , reduced forces on a unit test charge at  $r=a_2 > a_1$  and  $\theta=\theta_2$  are

$$Z_{\parallel} = - \int ds E_s(a_2, \theta_2, s; \omega) e^{-iks}, \quad Z_x = -i \int ds [E_x(a_2, \theta_2, s; \omega) - \beta G_y(a_2, \theta_2, s; \omega)] e^{-iks}.$$

Space-charge contributions for all multiples ( $\alpha_1=1$  or perfectly conducting at  $r=b^{(1)}$ ):

$$Z_{\parallel}^{\text{SC}} = \sum_{m=0}^{\infty} \frac{ikZ_0L \cos m\theta_2}{\pi\beta\gamma^2(1+\delta_{m0})} I_m(x_2) \mathcal{K}_m(x_1), \quad \mathcal{K}_m(x_i) = \left[ K_m(x_i) - \frac{K_m^{(1)}}{I_m^{(1)}} I_m(x_i) \right], \quad x_i = \frac{ka_i}{\gamma}$$

$$Z_x^{\text{SC}} = \sum_{m=0}^{\infty} \frac{ikZ_0L}{\pi\beta\gamma^3(1+\delta_{m0})} I_m(x_1) \left[ \cos \theta_2 \cos m\theta_2 \mathcal{K}_m(x_2) + \frac{m\gamma}{a_2} \sin \theta_2 \sin m\theta_2 \mathcal{K}'_m(x_2) \right]$$

The rest are from wall impedances. To any order  $a_1^{n_1} a_2^{n_2}$ , they are

$$Z_{\parallel}^{W,n_1,n_2} = -\frac{iL\mu_0\omega}{\pi\beta^2\gamma^2} \left( \frac{ka_1}{2\gamma} \right)^{n_1} \left( \frac{ka_2}{2\gamma} \right)^{n_2} \sum' \frac{\cos m\theta_2 \bar{\alpha}_1(m) K_m^{(1)}/I_m^{(1)}}{(1+\delta_{m0})(\frac{n_1+m}{2})!(\frac{n_1+m}{2})!(\frac{n_2-m}{2})!(\frac{n_2+m}{2})!}$$

$$Z_x^{W,n_1,n_2} = -\frac{iZ_0L}{\pi\beta\gamma^2 a_2} \left( \frac{ka_1}{2\gamma} \right)^{n_1} \left( \frac{ka_2}{2\gamma} \right)^{n_2} \sum' \frac{(n_2 \cos \theta_2 \cos m\theta_2 + m \sin \theta_2 \sin m\theta_2) \bar{\alpha}_1(m) K_m^{(1)}/I_m^{(1)}}{(1+\delta_{m0})(\frac{n_1-m}{2})!(\frac{n_1+m}{2})!(\frac{n_2-m}{2})!(\frac{n_2+m}{2})!}$$

where  $\bar{\alpha}_1(m) \equiv 1 - \alpha_1(m)$ ,  $\sum'$  implies from  $m=0$  to  $\min(n_1, n_2)$  with  $n_1-n_2$  and  $n_1-m$  even. The usual monopole and dipole pipe-wall impedances are

$$Z_0^{\parallel} = Z_{\parallel}^{W,0,0} = \frac{ikZ_0L \bar{\alpha}_1 K_0^{(1)}}{2\pi\beta\gamma^2 I_0^{(1)}} \quad \text{and} \quad Z_1^{\perp} = \frac{Z_x^{W,1,1}}{a_1} = \frac{iLZ_0k^2 \bar{\alpha}_1 K_1^{(1)}}{4\pi\beta\gamma^4 I_1^{(1)}}$$

**Multi-layer special cases:** [13] Pipe wall:  $b^{(1)} < r < b^{(2)} = b^{(1)} + t$ .

**Thin wall:** Good for low frequencies.  $t \rightarrow 0$  and  $E_s$  does not change across wall. At  $r=b^{(3)}$ , Case PC: perfectly conducting, Case PM: perfectly magnetic, and Case INF:  $b^{(2)} \rightarrow \infty$ .

$$\bar{\alpha}_1 = -\frac{\gamma^2\beta^2(1-\alpha_2) + 2ix\gamma\beta/m\zeta}{1 + \frac{x^2}{m^2} - \frac{ix\gamma\beta}{m} \left[ \frac{2}{\zeta(1-\alpha_2)} + \frac{\zeta(1-\alpha_2)}{2} \right]}, \quad \alpha_2^{\text{PC}} = -\alpha_2^{\text{PM}} = \frac{K_m(y)I_m(x)}{K_m(x)I_m(y)} \underset{m \neq 0}{\approx} \left( \frac{b^{(1)}}{b^{(3)}} \right)^2, \quad \alpha_2^{\text{INF}} = 0$$

$x = kb^{(1)}/\gamma, y = kb^{(3)}/\gamma, \zeta = Z_0\sigma_c t$ , and  $m \neq 0$ .

|   |  |
|---|--|
| <b>Thick wall:</b> Good for high frequencies. At $r = b^{(2)}$ , Case PC: perfectly conducting, Case PM: perfectly magnetic, and Case INF: $b^{(2)} \rightarrow \infty$ . For $m \geq 1$ ,  |  |
| $\bar{\alpha}_1 = \frac{-2\beta^2\gamma^2 \left[ 1 - \frac{(1+i)\Delta Q_\eta}{2m\beta\gamma^2} \right]}{1 - 2ip - \beta \left[ \frac{(1-i)Q_\alpha}{m\Delta} - \frac{(1+i)\Delta Q_\eta}{2m} \right] + \frac{Q_\alpha Q_\eta - m^2 p^2}{m^2 \gamma^2}}$ , $p = \frac{k^2 \delta^2}{2}$ , $\Delta = \mu_1 \beta k \delta_c$ , |  |
| $Q_\alpha = kb^{(1)} \frac{Q_2 - \alpha_2 P_2}{1 - \alpha_2}$ , $Q_\eta = kb^{(1)} \frac{Q_2 - \eta_2 P_2}{1 - \eta_2}$ , $Q_2 = \frac{K_m'^{(2)}}{K_m^{(2)}}$ , $P_2 = \frac{I_m'^{(2)}}{I_m^{(2)}}$   |  |
| Boundary conditions require $\alpha_2^{\text{PC}} = \eta_2^{\text{PM}} = \frac{K_m^{(2,3)} I_m^{(2)}}{I_m^{(2,3)} K_m^{(2)}}$ , $\eta_2^{\text{PC}} = \alpha_2^{\text{PM}} = \frac{K_m'^{(2,3)} I_m^{(2)}}{I_m'^{(2,3)} K_m^{(2)}}$ ,   |  |
| $\alpha_2^{\text{INF}} = \eta_2^{\text{INF}} = 0$ , with $I_m^{(p+1,p)} = I_m(\nu^{(p+1)} b^{(p)})$ , $I_m^{(p)} = I_m(\nu^{(p)} b^{(p)})$ and similar definitions for $I_m'$ , $K_m'$ , and $K_m'^{(2)}$ .   |  |

**Electric- and magnetic-dipole approximation:**  $\bar{\alpha}_1$  can also be derived [14] by approximating beam dipole motion as a superposition of oscillating electric and magnetic dipoles.

| Description   | Impedances  | Wakes   |
|---|---|---|
| <b>High frequency optical model:</b> [15] High frequency $k \gg 1/h$ , short-range $-z \ll h$ , transition length $L \ll kh^2$ , $h$ is minimum aperture. For tapered transition of angle $\theta$ , need $k \gg 1/h\theta$ . |   |   |
| Transitions, shallow cavities, collimators, irises:   | $Z^\parallel$ and $kZ^\perp$ are both constants similar for $kZ_d^\perp$ , $kZ_q^\perp$ , $W_d^\perp$ , $W_q^\perp$   | $W_\parallel = -Z^\parallel c\delta(z)$<br>$W_\perp = -kZ^\perp cH(-z)$   |
| (a) Axially symmetric examples: [16]-[18] (i) step-in transition (from $d$ to $b$ ), (ii) step-out (from $b$ to $d$ ), long collimator, shallow cavity with gap $g$ , (iii) thin iris   | (i) $Z_0^\parallel = Z_1^\perp = 0$ , (ii) $Z_0^\parallel = \frac{Z_0}{\pi} \ln \frac{d}{b}$ , $kZ_1^\perp = \frac{Z_0}{\pi} \left( \frac{1}{b^2} - \frac{1}{d^2} \right)$<br>(iii) $Z_0^\parallel = \frac{Z_0}{\pi} \ln \frac{d}{b}$ , $kZ_1^\perp = \frac{Z_0}{2\pi} \left( \frac{1}{b^2} - \frac{b^2}{d^4} \right)$  | where $b$ is small iris or pipe radius, $d$ is large pipe radius. Note: for shallow cavity, waves reflect from outer wall $\Rightarrow g \gtrsim k(d-b)^2$ ; for collimator, bottom length $\gg kb^2$ |
| (b) 3D, mirror symmetric in $x, y$ : [15] (i) flat step-out transition, aperture $2b$ to $2d$ , (ii) any step-in transition; iris with small (iii) flat (height $2b$ ), (iv) elliptical (axes $w$ by $b$ ), aperture          | (i) $kZ_y = \frac{\pi}{8} Z_0 \left( \frac{1}{b^2} - \frac{1}{d^2} \right)$ , $Z_q^y = \frac{1}{2} Z_d^y = \frac{1}{3} Z_y$<br>(ii) $Z_\parallel = Z_\perp = 0$ , (iii) $kZ_y = \frac{Z_0}{2\pi b^2}$ , $Z_q^y = Z_d^y = \frac{1}{2} Z_y$<br>(iv) $kZ_y = \frac{Z_0}{2\pi b^2}$ , $kZ_d^y = \frac{Z_0}{4\pi b^2} \left( 1 + \frac{b^2}{w^2} \right)$ , $kZ_q^y = \frac{Z_0}{4\pi b^2} \left( 1 - \frac{b^2}{w^2} \right)$ |   |
| <b>High frequency diffraction formulae:</b> $k \gg 1/b$<br>(a) Deep cavity (Fresnel diffraction)<br>[19, 1], cavity radius $d$ and gap $g$ .  | $Z_m^\parallel = \frac{\sqrt{2}Z_0}{(1+\delta_{m0})\pi^{3/2}b^{2m+1}} \sqrt{\frac{ig}{k}}$<br>$Z_1^\perp = \frac{2}{b^2 k} Z_0^\parallel$<br>Note: no reflections from outer wall $\Rightarrow g \lesssim k(d-b)^2$ .   | $W'_m = \frac{\sqrt{2}Z_0 c}{(1+\delta_{m0})\pi^{2}b^{2m+1}} \sqrt{\frac{g}{-z}}$<br>$W_1 = \frac{2}{b^2} \int W'_0 dz$   |

| Description   | Impedances  | Wakes   |
|---|---|---|
| (b) Periodic array of deep cavities (model for linear accelerator structures): [20]-[24] period $L$ , gap $g$ , outer cavity radius $d$ , with $g \lesssim k(d-b)^2$ .  | $Z_0^{\parallel} = \frac{iZ_0}{\pi kb^2} \left[ 1 + \frac{\alpha(g/L)L}{b} \sqrt{\frac{2\pi i}{kg}} \right]^{-1}$<br>$\alpha(\zeta) \approx 1 - 0.465\sqrt{\zeta} - 0.070\zeta$<br>$Z_1^{\perp} = \frac{2}{b^2k} Z_0^{\parallel}$   | $\frac{W'_0}{L} = \frac{Z_0c}{\pi b^2} e^{\eta(z)^2} \operatorname{erfc}[\eta(z)]$<br>$\eta(z) = \frac{\alpha L}{b} \sqrt{\frac{-2\pi z}{g}}$<br>$W_1 = \frac{2}{b^2} \int W'_0 dz$   |
| Numerical fit:[24, 25] valid over larger $z$ range: $-z/L \leq 0.15$ , $0.34 \leq b/L \leq 0.69$ , $0.54 \leq g/L \leq 0.89$ .  | $W'_0 = \frac{Z_0c}{\pi b^2} \exp\left(-\sqrt{\frac{z}{z_0}}\right)$ , $W_1 = \frac{4Z_0cz_1}{\pi b^4} \left[ 1 - \left( 1 + \sqrt{\frac{z}{z_1}} \right) \exp\left(-\sqrt{\frac{z}{z_1}}\right) \right]$<br>$z_0 = -0.41 \frac{b^{1.8}g^{1.6}}{L^{2.4}}$ , $z_1 = -0.17 \frac{b^{1.79}g^{0.38}}{L^{1.17}}$                       |   |
| <b>Bethe's dipole moments</b> of a hole of radius $a$ on beam pipe wall [26].   | Electric and magnetic dipole moments when wavelength $\gg a$ :<br>$\vec{d} = -\frac{2\epsilon_0}{3}a^3\vec{E}$ , $\vec{m} = -\frac{4}{3\mu_0}a^3\vec{B}$<br>$\vec{E}$ and $\vec{B}$ are electric and magnetic flux density at hole when hole is absent. This is a diffraction solution for a thin-wall pipe.                      |   |
| <b>Small 3D obstacle</b> on beam pipe: [27, 28] size $\ll b$ , low freq. $k \ll 1/(\text{size})$ ; $\phi$ azimuthal angular position of object.   | $Z_0^{\parallel} = -ikc\mathcal{L}$ ,<br>$Z_1^{\perp} = \frac{4}{b^2k} Z_0^{\parallel} \cos \phi$   | $W'_0 = -c^2 \mathcal{L} \delta'(z)$<br>$W_1 = \frac{4}{b^2} \int W'_0 \cos \phi dz$  |
| Elliptical hole: major and minor radii are $a$ and $d$ . $K(m)$ and $E(m)$ are complete elliptical functions of the first and second kind, with $m = 1 - m_1$ and $m_1 = (d/a)^2$ . For long ellipse perpendicular to beam, major axis $a \ll b$ , beam pipe radius, because the curvature of the beam pipe has been neglected here [29]. | $\alpha_e + \alpha_m = \begin{cases} \frac{\pi a^3 m_1^2 [K(m) - E(m)]}{3E(m)[E(m) - m_1 K(m)]} & \parallel \text{beam} \\ \frac{\pi a^3 [E(m) - m_1 K(m)]}{3[K(m) - E(m)]} & \text{long ellipse} \end{cases}$<br>$\alpha_e + \alpha_m \xrightarrow[m \rightarrow 0]{\text{circular}} \frac{2a^3}{3}$ circular hole $a = d \ll b$ | $\frac{\pi d^4 [\ln(4a/d) - 1]}{3a} \quad d \ll b$<br>$\frac{\pi a^3}{3 [\ln(4a/d) - 1]} \quad a \ll b$<br>Above are for $t \ll a$ . When $t \geq a$ , $\times 0.56$ when hole is circular and $\times 0.59$ when hole is long-elliptic.<br>For higher frequency correction, add to $\alpha_e + \alpha_m$ the extra term,<br>$+ \frac{2\pi a^3}{3} \left[ \frac{11k^2 a^2}{30} \right] \text{circular}, \begin{cases} -\frac{\pi ad^2}{3} \left[ \frac{k^2 a^2}{5} \right] & \parallel \text{beam} \\ +\frac{2\pi a^3}{3} \left[ \frac{2k^2 a^2}{5 [\ln(4a/d) - 1]} \right] & \perp \text{beam} \end{cases}$ long ellipse |
| Rectangular slot: length $L$ , width $w$ .  | $\alpha_e + \alpha_m = w^3 (0.1814 - 0.0344w/L) \quad t \ll a, \quad \times 0.59 \text{ when } t \geq a$  |   |
| Rounded-end slot: length $L$ , width $w$ .  | $\alpha_e + \alpha_m = w^3 (0.1334 - 0.0500w/L) \quad t \ll a, \quad \times 0.59 \text{ when } t \geq a$  |   |

| Description   | Impedances   | wake  |
|---|--|---|
| Annular-ring-shaped cut: inner and outer radii $a$ and $d = a + w$ with $w \ll d$ .   | $\alpha_e + \alpha_m = \frac{\pi^2 d^2 a}{2 \ln(32d/w) - 4} - \frac{\pi^2 w^2(a+d)}{16}$ $t \ll d$<br>$\alpha_e + \alpha_m = \pi d^2 w - \frac{1}{2} w^2(a+d)$ $t \geq d$  |   |
| Half ellipsoidal protrusion with semi axes $h$ radially, $a$ longitudinally, and $d$ azimuthally. ${}_2F_1$ is the hypergeometric function.   | $\alpha_e + \alpha_m = 2\pi a h d \left[ \frac{1}{I_b} + \frac{1}{I_c - 3} \right]$<br>$I_b = {}_2F_1\left(1, 1; \frac{5}{2}; 1 - \frac{h^2}{a^2}\right), \quad I_c = {}_2F_1\left(1, \frac{1}{2}; \frac{5}{2}; 1 - \frac{a^2}{h^2}\right), \quad \text{if } a = d$<br>$\alpha_e + \alpha_m = \pi a^3 \quad \text{if } a = d = h, \quad \frac{2\pi h^3}{3[\ln(2h/a) - 1]} \quad \text{if } a = d \ll h$<br>$\alpha_e + \alpha_m = \frac{8h^3}{3} \left[ 1 + \left( \frac{4}{\pi} - \frac{\pi}{4} \right) \frac{a}{h} \right] \quad \text{if } a \ll h = d$<br>$\alpha_e + \alpha_m = \frac{8\pi h^4}{3a} \left[ \ln \frac{2a}{h} - 1 \right] \quad \text{if } a \gg h = d$ |   |
| <b>Small inductive objects-2D:</b> [27, 30]<br>small cavities, shallow irises, and transitions at low freq. ( $h \ll b, k \ll 1/h$ ); $h$ is height of object, $g$ is gap of cavity or length of iris; $\mathcal{L}$ is inductance. For tapered transition pair: $\theta$ is taper angle. | $Z_0^\parallel = -ikc\mathcal{L}, \quad Z_1^\perp = \frac{2}{b^2 k} Z_0^\parallel$   | $W'_0 = -c^2 \mathcal{L} \delta'(z), \quad W_1 = \frac{2}{b^2} \int W'_0 dz$<br><br>Pill box $g \lesssim h \ll b : \mathcal{L} = \frac{Z_0}{2\pi cb} \left[ gh - \frac{g^2}{2\pi} \right]$<br>Shallow iris $g \lesssim h \ll b : \mathcal{L} = \frac{Z_0 h^2}{4cb}$<br>Transition pair $g \gg h, h \ll b : \mathcal{L} = \frac{Z_0 h^2}{2\pi^2 cb} \left( \ln \frac{2\pi b}{h} + \frac{1}{2} \right)$<br>Tapered : $\mathcal{L} = \frac{Z_0 h^2}{\pi^2 cb} \left[ \ln \left( \frac{b\theta}{h} - 2\theta \cot \theta \right) + \frac{3}{2} - \gamma_e - \psi \left( \frac{\theta}{\pi} \right) - \frac{\pi}{2} \cot \theta - \frac{\pi}{2\theta} \right]$<br>$\gamma_e \approx 0.57721$ is Euler's constant, $\psi(x)$ is psi function. |
| <b>Wall roughness inductive model:</b> [35]<br>1-D axisymmetric bump on beam pipe, $h(z)$ or 2-D bump $h(z, \theta)$ . Valid for low frequency $k \ll (\text{bump length or width})^{-1}$ , $h \ll b$ , and $ \nabla h  \ll 1$ . See also [36]  | 1-D:<br>$Z_0^\parallel = -\frac{2ikZ_0}{b} \int_0^\infty \kappa  \tilde{h}(\kappa) ^2 d\kappa$<br>with spectrum $\tilde{h}(k) = \frac{1}{2\pi} \int_{-\infty}^\infty h(z) e^{-ikz} dz$<br><br>2-D:<br>$Z_0^\parallel = -\frac{4ikZ_0}{b} \sum_{m=-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{\kappa^2}{\sqrt{\kappa^2 + m^2/b^2}}  \tilde{h}_m(\kappa) ^2 d\kappa$<br>with spectrum $\tilde{h}_m(k) = \frac{1}{(2\pi)^2} \int_0^{2\pi} d\theta \int_{-\infty}^{\infty} dz h(z, \theta) e^{-ikz - im\theta}$<br>Note: small periodic corrugations model is also used for wall roughness impedance estimation.   |   |
| <b>Small periodic corrugations:</b> (a) [31, 32] $L \lesssim h \ll b$ , $k \ll 1/h$ ; $L$ period, $h$ depth, $g$ gap, $\wp$ principal value; $\beta_g c$ group velocity.<br>(b) [33] $L \gg h, L \ll b$ , $k \ll 1/h$ ; $k_L = 2\pi/L$ .  | $\frac{Z_0^\parallel}{L} = \frac{Z_0}{\pi b^2} \left[ \pi k_r \delta(k^2 - k_r^2) + i \wp \left( \frac{k}{k^2 - k_r^2} \right) \right], \quad \frac{W'_0}{L} = \frac{Z_0 c}{\pi b^2} \cos k_r z$<br>$Z_1^\perp = \frac{2}{b^2 k} Z_0^\parallel, \quad k_r = \sqrt{\frac{2L}{bgh}}; \quad (1 - \beta_g) = \frac{4hg}{bL}, \quad W_1 = \frac{2}{b^2} \int W'_0 dz$<br><br>$\frac{Z_0^\parallel}{L} = \frac{Z_0 h^2 k_L^{3/2}}{8\pi b} (-ik)^{1/2}$   | $\frac{W'_0}{L} = -\frac{Z_0 c h^2 k_L^3}{16\pi^{3/2} b} \frac{1}{(-k_L z)^{3/2}}$  |

| Description   | Impedances   | Wakes  |
|---|--|--|
| <b>Thin dielectric or ferrite layer on pipe:</b> [34] thickness $h \ll b$ .   |  | Like small periodic corrugations (a), but $k_r = \left[ \frac{2\epsilon_r}{(\epsilon_r\mu_r - 1)bh} \right]^{1/2}$ , with relative dielectric constant $\epsilon_r$ and magnetic permeability $\mu_r$ .  |
| <b>Coherent synchrotron radiation (CSR):</b> [37, 38, 39]<br>Bunch moves in free space on a circle of radius $R$ ; $k \ll \gamma^3/R$ . See Sec.2.4.15.                                 | $Z_0^\parallel = \frac{Z_0}{2 \cdot 3^{1/3} \pi} \Gamma\left(\frac{2}{3}\right) \left[\frac{ik}{R^2}\right]^{1/3}$   | $W_0' = -\frac{Z_0 c}{2 \cdot 3^{4/3} \pi R^{2/3}} \frac{1}{z^{4/3}}$  |
|   |  | $\Gamma(2/3) \approx 1.3541$ . Note: non-zero wake for test particle <i>ahead</i> of driving particle. $W_0'(0^+)/L \approx 0.1 Z_0 c \gamma^4 / R^2$ . This is also used to approximate effect at high $k$ for beam in beam pipe; shielded (suppressed) for $k \lesssim R^{1/2} b^{-3/2}$ . |
| <b>Round collimator:</b><br>(a) [40] low frequency $k \ll 1/d$ .  | $Z_1^\perp = -0.3i \frac{Z_0}{d}$<br>collimator radius $d \ll b$ .   | $W_1 = -0.3 \frac{Z_0 c}{d} \delta(z)$<br>collimator radius $d \ll b$ .  |
| (b) High frequency $k \gg 1/d$ ; if tapered, angle $\theta \gg 1/(kd)$ .  |  | See optical model formulae (a) above   |
| (c) [41] For any frequency, small angle, $d'(s) \ll 1$ , $kdd' \ll 1$ , with $d(s)$ pipe profile versus longitudinal position $s$ , and $d'$ is derivative of $d$ with respect to $s$ . | $Z_0^\parallel = \frac{-iZ_0 k}{4\pi} \int ds (d')^2$<br>$Z_1^\perp = \frac{-iZ_0}{2\pi} \int ds \left(\frac{d'}{d}\right)^2$<br>$\Rightarrow$ symm. tapers of angle $\theta \ll 1$ :<br>$Z_1^\perp = \frac{-iZ_0}{\pi} \theta \left(\frac{1}{d} - \frac{1}{b}\right)$ | $W_0' = \frac{Z_0 c}{4\pi} \int ds (d')^2 \delta'(z)$<br>$W_1 = -\frac{Z_0 c}{2\pi} \int ds \left(\frac{d'}{d}\right)^2 \delta(z)$<br>$W_1 = -\frac{Z_0 c}{\pi} \theta \left(\frac{1}{d} - \frac{1}{b}\right) \delta(z)$   |
| <b>Flat collimator:</b> [42]<br>low frequency, small angle, $h'(s) \ll 1$ , $h \ll w \ll \ell$ , with $h(s)$ vertical profile, $w$ width, $\ell$ length                                 | $Z_y = \frac{-iZ_0 w}{4} \int ds \frac{(h')^2}{h^3}$   | $W_y = -\frac{Z_0 c w}{4} \int ds \frac{(h')^2}{h^3} \delta(z)$  |
| <b>Pill-box cavity —low frequency:</b> [43] cavity radius $d$ , gap $g$ ; $S = d/b$ . When $g \gg 2(d-b)$ , replace $g$ by $d-b$ . Valid for $k \ll 1/d$ .                              | $Z_0^\parallel = -ik \frac{Z_0 g}{2\pi} \ln S$<br>$Z_1^\perp = -i \frac{Z_0 g}{\pi b^2} \frac{S^2 - 1}{S^2 + 1}$   | $W_0' = -\frac{Z_0 c g}{2\pi} \ln S \delta'(z)$<br>$W_1 = -\frac{Z_0 c g}{\pi b^2} \frac{S^2 - 1}{S^2 + 1} \delta(z)$  |
|   | Effect will be one half for a step in the beam pipe from radius $b$ to radius $d$ , or vice versa, when $g \gg 2(d-b)$ .   |  |
| <b>Resonator model:</b><br>[1] for $m$ -th azimuthal mode, with shunt impedance $R_s^{(m)}$ , quality factor $Q$ , and resonant frequency $k_r$ .                                       | $Z_m^\parallel = \frac{R_s^{(m)}}{1 + iQ(k_r/k - k/k_r)}$<br>$Z_m^\perp = \frac{R_s^{(m)}/k}{1 + iQ(k_r/k - k/k_r)}$   | $W_m = \frac{R_s^{(m)} c k_r}{Q k_r} e^{\alpha z} \sin \bar{k}_r z$<br>where $\alpha = k_r/(2Q)$<br>$\bar{k}_r = \sqrt{ k_r^2 - \alpha^2 }$  |
|   | Valid only close to $k_r$ . As $k \rightarrow \infty$ , $Z_0^\parallel \rightarrow k^{-1/2}$ for non-periodic cavities and $\rightarrow k^{-3/2}$ for an infinite array of cavities. [16, 46]  |  |

| Description   | Impedances   | Wakes   |
|---|--|---|
| <b>Closed pill-box cavity:</b> [44] resonant frequencies $k_{mnp}$ and “circuit” $(R_s/Q)_{mnp}$ [45], where $m, n, p$ , are azimuthal, radial, longitudinal mode numbers. Cavity radius $d$ and length $g$ ; $x_{mn}$ is $n^{\text{th}}$ zero of Bessel function $J_m$ . | $k_{mnp}^2 = \frac{x_{mn}^2}{d^2} + \frac{p^2\pi^2}{g^2}$ $\left[ \frac{R_s}{Q} \right]_{0np} = \frac{Z_0}{x_{0n}^2 J_0'^2(x_{0n})} \frac{8}{\pi g k_{0np}} \begin{cases} \sin^2 \frac{g k_{0np}}{2\beta} \times \frac{1}{1 + \delta_{0p}} & p \text{ even} \\ \cos^2 \frac{g k_{0np}}{2\beta} & p \text{ odd} \end{cases}$ $\left[ \frac{R_s}{Q} \right]_{1np} = \frac{Z_0}{J_1'^2(x_{1n})} \frac{2}{\pi g d^2 k_{1np}^2} \begin{cases} \sin^2 \frac{g k_{1np}}{2\beta} & p \neq 0 \text{ and even} \\ \cos^2 \frac{g k_{1np}}{2\beta} & p \text{ odd} \end{cases}$ |   |
| <b>Curvature impedance:</b> [47] Smooth toroidal chamber of rectangular cross section, width $b - a$ , height $h$ , inner radius $a$ , outer radius $b$ , and $R = \frac{1}{2}(a + b)$ . As Lorentz factor $\gamma \rightarrow \infty$ , a contribution remains.          | Valid from zero frequency up to just below synchronous resonant modes, i.e., $0 < \nu < \sqrt{R/h}$ with $\nu = kh$ ,  | $Z_0^{\parallel} = \frac{ikZ_0h^2}{\pi^2 R} \left\{ \left[ 1 - e^{-2\pi(b-R)/h} - e^{-2\pi(R-a)/h} \right] \left[ 1 - 3 \left( \frac{\nu}{\pi} \right)^2 \right] + 0.05179 - 0.01355 \left( \frac{\nu}{\pi} \right)^2 \right\} + \rho kR$ $\approx \frac{ikZ_0h^2}{\pi^2 R} \left[ A - 3B \left( \frac{\nu}{\pi} \right)^2 \right].$ <p>where <math>\rho</math> is quadratic in <math>\nu</math>. As <math>(b-a)/h</math> increases, <math>\rho</math> vanishes exponentially and <math>A \approx B \approx 1</math>. In general, <math>A/B \approx 1</math> implying <math>\text{Im}Z_0^{\parallel}</math> changes sign (a node) near <math>\nu = \pi/\sqrt{3}</math>.</p> |
| <b>Kicker with window-frame magnet:</b> [49] width $a$ , height $b$ , length $L$ , beam offset $x_0$ horizontally, and all image current carried by conducting current plates.  | $Z_0^{\parallel} = \frac{k^2 c^2 \mu_0^2 L^2 x_0^2}{4a^2 Z_k}$ $Z_1^{\perp} = \frac{kc^2 \mu_0^2 L^2}{4a^2 Z_k}$   | $W'_0 = -\frac{c^3 \mu_0^2 L^2 x_0^2}{4a^2 Z_k} \delta''_0(z)$ $W_1 = -\frac{c^3 \mu_0^2 L^2}{4a^2 Z_k} \delta'(z)$   |
|   | $Z_k = -ikc\mathcal{L} + Z_g$ with $\mathcal{L} \approx \mu_0 b L/a$ the inductance of the windings and $Z_g$ the impedance of the generator and the cable. If the kicker is of C-type magnet, $x_0$ in $Z_0^{\parallel}$ should be replaced by $(x_0 + b)$ .  |   |
| <b>Traveling-wave kicker</b> [49] with characteristic impedance $Z_c$ for the cable, and a window magnet of width $a$ , height $b$ , and length $L$ . Valid for frequency below cutoff.   | $Z_0^{\parallel} = \frac{Z_c}{4} \left[ 2 \sin^2 \frac{\theta}{2} - i \sin \theta \right]$ $Z_1^{\perp} = \frac{Z_c L}{4ab} \left[ \frac{1 - \cos \theta}{\theta} - i \frac{\sin \theta}{\theta} \right]$  | $W'_0 = \frac{Z_c c}{4} \left[ \delta(z) - \delta \left( z + \frac{L}{\beta_{\text{ph}}} \right) \right]$ $W_1 = \frac{Z_c \beta c}{4ab} \left[ H(z) - H \left( z + \frac{L}{\beta_{\text{ph}}} \right) \right]$  |
|   | $\theta = kL/\beta_{\text{ph}}$ denotes the electrical length of the kicker windings and $\beta_{\text{ph}}c = Z_{cac}/(Z_0b)$ is the matched transmission-line phase velocity of the capacitance-loaded windings. Here, $\beta_{\text{ph}} \ll \beta \rightarrow 1$ , the beam velocity.  |   |

| Description  | Impedances  | Wakes   |   |  |  |
|--|---|---|---|--|--|
| <b>Strip-line BPMs</b><br>(pair): [48] length $L$ , angle each subtending to pipe axis $\phi_0$ , forming transmission lines of characteristic impedance $Z_c$ with pipe.  | $Z_0^{\parallel} = 2Z_c \left[ \frac{\phi_0}{2\pi} \right]^2 [2 \sin^2 kL - i \sin 2kL]$<br>$Z_1^{\perp} = \left[ \frac{Z_0^{\parallel}}{k} \right]_{\text{pair}} \frac{1}{b^2} \left[ \frac{4}{\phi_0} \right]^2 \sin^2 \frac{\phi_0}{2}$  | $W'_0 = 2Z_c c \left[ \frac{\phi_0}{2\pi} \right]^2 [\delta(z) - \delta(z+2L)]$   | $W_1 = \frac{8Z_c c}{\pi^2 b^2} \sin^2 \frac{\phi_0}{2} [H(z) - H(z+2L)]$                                   |  |  |
| The strip-lines are assumed to terminate with impedance $Z_c$ at the upstream end.   |   |   |   |  |  |
| <b>Wakes for a Gaussian Bunch:</b>   |   |   |   |  |  |
| <p>The <i>bunch wakes</i> of a bunch with longitudinal charge distribution <math>\lambda_z</math>, are given by <math>\mathcal{W}'_m(z) = \int_{-\infty}^0 W_m(x) \lambda_z(z-x) dx</math>, <math>\mathcal{W}_m(z) = \int_{-\infty}^0 W_m(x) \lambda_z(z-x) dx</math>. In the following we give bunch wakes of a Gaussian bunch [<math>\lambda_z = e^{-(z/\sigma_z)^2/2}/(\sqrt{2\pi}\sigma_z)</math>, with <math>\sigma_z</math> the rms bunch length] for wakefield forms found in the tables above, and also give their first moments <math>\langle \mathcal{W} \rangle = \int_{-\infty}^{\infty} \mathcal{W}(z) \lambda_z(z) dz</math> and the rms <math>\mathcal{W}_{\text{rms}} = \sqrt{\langle \mathcal{W}^2 \rangle - \langle \mathcal{W} \rangle^2}</math>. Here the <math>z</math> dependence alone is considered and the wake coefficient is scaled out; for a specific problem, the appropriate coefficients, found in the tables above, need to be included at the end.</p> |   |   |   |  |  |
| <p>Note: for power law wakes with <math>-2 &lt; \alpha &lt; -1</math>, <math>\mathcal{W}</math> is obtained using integration by parts [38]. It is assumed that in the range <math> z  \ll \sigma_z</math> the wake form changes so that <math>\int_{-\infty}^{\infty} W(z) dz = 0</math>. Consequently, <math>\mathcal{W}</math> can be obtained without knowing the details of <math>W</math> at very short range.</p>   |   |   |   |  |  |
| Wake form, $W$   | Bunch wake, $\mathcal{W}$   | $\langle \mathcal{W} \rangle$   | $\mathcal{W}_{\text{rms}}$  |  |  |
| <b>Circuit Models:</b><br>Resistive: $\delta(z)$<br>Inductive: $\delta'(-z)$<br>Capacitive: $H(-z)$  | $\frac{1}{\sqrt{2\pi}\sigma_z} e^{-(z/\sigma_z)^2/2}$<br>$\frac{z}{\sqrt{2\pi}\sigma_z^3} e^{-(z/\sigma_z)^2/2}$<br>$\frac{1}{2} \left[ 1 + \text{erf} \left( \frac{-z}{\sqrt{2}\sigma_z} \right) \right]$  | $\frac{1}{2\sqrt{\pi}\sigma_z}$<br>0<br>$\frac{1}{2}$   | $\frac{0.111}{\sigma_z}$<br>$\frac{1}{\sqrt{6\pi} 3^{1/4} \sigma_z^2}$<br>$\frac{1}{2\sqrt{3}}$             |  |  |
| <b>Power Law:</b> $(-z)^\alpha$<br>Low freq. resistive wall ( $W_m$ ) and Fresnel diffraction ( $W'_m$ ): $\alpha = -\frac{1}{2}$<br>Fresnel diffraction ( $W_m$ ): $\alpha = \frac{1}{2}$<br>Low freq. resistive wall ( $W'_m$ ) and small periodic corrugations ( $W'_0$ ): [50] $\alpha = -\frac{3}{2}$<br>CSR ( $W'_0$ ): $z^\alpha$ with $\alpha = -\frac{4}{3}$ (note: $z > 0$ )   | $f(-z/\sigma_z) \sigma_z^\alpha$ , with $f(x)$ given by (upper/lower sign for $x \gtrless 0$ ):<br>$\sqrt{\frac{\pi x }{8}} e^{-x^2/4} [I_{-1/4} \pm I_{1/4}] \Big _{x^2/4}$<br>$\sqrt{\frac{\pi}{32}} \int_{-\infty}^x  y ^{1/2} e^{-y^2/4} [I_{-1/4} \pm I_{1/4}] \Big _{y^2/4} dy$<br>$\sqrt{\frac{\pi x ^3}{8}} e^{-x^2/4} [I_{1/4} - I_{-3/4} \pm I_{-1/4} \mp I_{3/4}] \Big _{x^2/4}$<br>$-\frac{3}{\sqrt{2\pi}} \int_0^{\infty} \frac{(x+y)e^{-(x+y)^2/2}}{y^{1/3}} dy$  | $\frac{0.723}{\sqrt{\sigma_z}}$<br>$0.489\sqrt{\sigma_z}$<br>$-\frac{0.489}{\sigma_z^{3/2}}$<br>$-\frac{0.758}{\sigma_z^{4/3}}$ | $\frac{0.292}{\sqrt{\sigma_z}}$<br>$0.374\sqrt{\sigma_z}$<br>$0.516\sigma_z^{3/2}$<br>$0.532\sigma_z^{4/3}$ |  |  |
| <b>Resonator Model:</b><br>$\begin{cases} \sin(-k_r z) \\ \cos(k_r z) \end{cases} e^{\alpha_r z}$ , with $k_r, \alpha_r$ , real  | $\mathcal{W} = f(-z/\sigma_z), \text{ with } f(x) = \frac{1}{2} e^{-(k_r^2 - \alpha_r^2)\sigma_z^2/2 - \alpha_r \sigma_z x} \times \begin{cases} \text{Im} \\ \text{Re} \end{cases} e^{ik_r \sigma_z (x - \alpha_r \sigma_z)} \left\{ 1 + \text{erf} \left[ \frac{(ik_r - \alpha_r)\sigma_z + x}{\sqrt{2}} \right] \right\}$<br>$\langle \mathcal{W} \rangle = \frac{1}{2} e^{-(k_r^2 - \alpha_r^2)\sigma_z^2} \begin{cases} \text{Im} \\ \text{Re} \end{cases} e^{-i2k_r \alpha_r \sigma_z^2} \left\{ 1 + \text{erf} [(ik_r - \alpha_r)\sigma_z] \right\}$ |   |   |  |  |

### Sec.3.2: IMPEDANCES AND WAKE FUNCTIONS

- [19] J. Lawson, Rutherford Report RHEL/M 144 (1968); K. Bane, M. Sands, SLAC-PUB-4441 (1987)
- [20] R. Gluckstern, PR D39 (1989) 2773, 2780
- [21] G. Stupakov, PAC 95, 3303
- [22] K. Yokoya, K. Bane, PAC 99, 1725
- [23] A. Fedotov, R. Gluckstern, M. Venturini, PRST-AB 2 (1999) 064401
- [24] K. Bane et al, ICAP 98, 137
- [25] K. Bane, SLAC-PUB (2003) 9663
- [26] M. Sands, SLAC note PEP-253 (1977); H.A. Bethe, PR 66 (1944) 163
- [27] S.S. Kurennoy, PR E55 (1997) 3529; S. Kurennoy, R. Gluckstern, *ibid* (1997) 3533
- [28] S.S. Kurennoy, PA 39 (1992) 1; PA 50 (1995) 167; R.L. Gluckstern, PR A46, 1106, 1110 (1992); S.S. Kurennoy, R.L. Gluckstern, G.V. Stupakov, PR E52 (1995) 4354
- [29] A. Fedotov, PhD Thesis, U. Maryland (1997)
- [30] S.S. Kurennoy, G.V. Stupakov, PA 45 (1994) 95
- [31] A. Novokhatski, A. Mosnier, PAC 97, 1661
- [32] K. Bane, A. Novokhatski, SLAC-AP-117 (1999)
- [33] G. Stupakov, Proc. 19th Advanced ICFA Beam Dynamics Workshop (Arcidosso, 2000), 141
- [34] K.Y. Ng, PR D42 (1990) 1819; A. Burov, A. Novokhatski, HEACC 92, 537
- [35] G.V. Stupakov, PRST-AB 1, 064401 (1998)
- [36] K. Bane et al, PAC 97, 1738
- [37] J.B. Murphy et al, PAC 95, 2980; PA 57 (1997) 9
- [38] Y. Derbenev et al, DESY-TESLA-FEL 95-05 (1995)
- [39] R. Warnock, P. Morton, SLAC-PUB-4562 (1988); R. Warnock, Proc. 4th Advanced ICFA Beam Dynamics Workshop (1990) 151
- [40] K. Bane, P. Morton, LINAC 86, 490
- [41] K. Yokoya, CERN-SL-90-88-AP (1988)
- [42] G.V. Stupakov, SLAC-PUB-7167 (1996)
- [43] E. Keil, B. Zotter, PA 3 (1972) 11; K.Y. Ng, Fermilab Report FN-389 (1981)
- [44] T. Weiland and B. Zotter, PA 11 (1981) 143
- [45] G. Dome, PAC 85, 2531
- [46] S.A. Heifets, S.A. Kheifets, PA 25 (1990) 61; A. Hofmann, T. Risselada, B. Zotter, Proc. 4th ICFA Beam Dynamics workshop (1990) 138
- [47] K.Y. Ng, R. Warnock, PAC 89, 798; PR D40 (1989) 231
- [48] K.Y. Ng, PA 23 (1988) 93
- [49] T. Toyama, et al, HB 2006, 140; G. Nassibian, CERN/PS 84-25 (BR) (1984); CERN 85-68 (BR) (1986)
- [50] A. Piwinski, DESY Report 72/72 (1972)

### 3.2.5 Effective Impedance [1, 2, 3]

*T. Suzuki, KEK*

The effective impedance is used to estimate complex coherent tune shifts approximately. Analytic sums are found for a sinusoidal mode [4] and approximately for a Gaussian bunch [5, 6] in the case of a resonator impedance.

#### Definition longitudinal

$$\left(\frac{Z_{\parallel 0}}{\omega}\right)_{\text{eff}}^{[m]} = \frac{\sum_{p=-\infty}^{\infty} \frac{Z_{\parallel 0}(\omega_p)}{\omega_p} h_m(\omega_p)}{\sum_{p=-\infty}^{\infty} h_m(\omega_p)} \quad (1)$$

$$\omega_p = \omega_0(pn_B + \mu + m\nu_s)$$

$$\mu = \text{coupled bunch mode number}$$

$$m = \text{azimuthal mode number}$$

#### transverse

$$(Z_{\perp 1})_{\text{eff}}^{[m]} = \frac{\sum_{p=-\infty}^{\infty} Z_{\perp 1}(\omega_p) h_m(\omega_p - \omega_{\xi})}{\sum_{p=-\infty}^{\infty} h_m(\omega_p - \omega_{\xi})} \quad (2)$$

$$\omega_p = \omega_0(pn_B + \mu + \nu_{\beta} + m\nu_s)$$

$$\eta = \alpha - \frac{1}{\gamma^2}, \quad \nu_{\beta} \text{ betatron tune}$$

$$\omega_{\xi} = \frac{\xi}{\eta} \nu_{\beta} \omega_0, \quad \xi = \frac{\Delta \nu_{\beta}}{\nu_{\beta}} / \frac{\Delta p}{p}$$

Tab.1 gives the spectral density  $h_m(\omega) \propto |\tilde{\lambda}_m(\omega)|^2$ , where  $\tilde{\lambda}_m(\omega)$  is the Fourier transform of the line density,  $\tau$  is the full bunch length in time,  $\sigma_{\tau}$  is the rms bunch length in time.  $h_m$  is chosen to be dimensionless. See also Tabs.2 & 3.

Table 1: Spectral densities  $h_m(\omega)$  for various modes.

| mode       | $h_m(\omega)$   |
|------------|---|
| sinusoidal | $(m+1)^2 \frac{1 + (-1)^m \cos \omega \tau}{[(\frac{\omega \tau}{\pi})^2 - (m+1)^2]^2}$ |
| Legendre   | $\frac{[J_{m+1/2}(\omega \tau/2)]^2}{ \omega \tau/2 }$                                  |
| Hermite    | $(\omega \sigma_{\tau})^{2m} e^{-\omega^2 \sigma_{\tau}^2}$                             |

Table 2: Complex longitudinal frequency shifts.  $\omega_s = \nu_s \omega_0$ ,  $\beta = v/c$ .

| mode                            | $\Omega_m - m\omega_s$   |
|---------------------------------|--|
| sinusoidal<br>(parabolic bunch) | $\frac{im}{m+1} \times$<br>$\frac{12N_B e^2 \eta}{T_0 E \beta^2 \omega_s \tau^3} \left( \frac{Z_{  0}}{\omega} \right)_{\text{eff}}^{[m]}$   |
| Legendre<br>(parabolic bunch)   | $\frac{12i \Gamma(m + \frac{1}{2})}{\sqrt{\pi} (m-1)!} \times$<br>$\frac{N_B e^2 \eta}{T_0 E \beta^2 \omega_s \tau^3} \left( \frac{Z_{  0}}{\omega} \right)_{\text{eff}}^{[m]}$    |
| Hermite<br>(Gaussian bunch)     | $\frac{i \Gamma(m + \frac{1}{2})}{2\pi 2^m (m-1)!} \times$<br>$\frac{N_B e^2 \eta}{T_0 E \beta^2 \omega_s \sigma_\tau^3} \left( \frac{Z_{  0}}{\omega} \right)_{\text{eff}}^{[m]}$ |

Table 3: Complex transverse frequency shifts for various beam models.  $\omega_\beta = \nu_\beta \omega_0$ .

| model                        | $\Omega_m - \omega_\beta - m\omega_s$   |
|------------------------------|---|
| sinusoidal<br>(square bunch) | $-\frac{i}{1+m} \times$<br>$\frac{N_B e^2 c}{2\omega_\beta T_0 E \tau} (Z_{\perp1})_{\text{eff}}^{[m]}$   |
| Legendre<br>(square bunch)   | $-\frac{i}{2\sqrt{\pi}} \frac{\Gamma(m + \frac{1}{2})}{m!} \times$<br>$\frac{N_B e^2 c}{\omega_\beta T_0 E \tau} (Z_{\perp1})_{\text{eff}}^{[m]}$     |
| Hermite<br>(Gaussian bunch)  | $-\frac{i}{4\pi} \frac{\Gamma(m + \frac{1}{2})}{2^m m!} \times$<br>$\frac{N_B e^2 c}{\omega_\beta T_0 E \sigma_\tau} (Z_{\perp1})_{\text{eff}}^{[m]}$ |

## References

- [1] A.W. Chao, Physics of Collective Beam Instabilities in High Energy Accelerators, Wiley (1993)
- [2] F.J. Sacherer, PAC 77, 1393
- [3] F.J. Sacherer, 9th Int.Conf.High Energy Acc. (1974) 347
- [4] B. Zotter, CERN/ISR-TH/78-16 (1978)
- [5] B. Zotter, CERN/ISR-TH/80-03 (1980)
- [6] K. Balewski, R.D. Kohaupt, DESY 90-152 (1990)

## 3.2.6 Parasitic Loss

P. Wilson, SLAC, B. Zotter, CERN  
Y.-H. Chin, KEK

**Total loss factor** The energy change (parasitic loss) of a bunch of charge  $q$  and normalized line density  $\lambda(t)$ , traversing a structure with longitudinal impedance  $Z_{||}$  (Sec.3.2) is

$$\Delta E = -\kappa_{||} q^2 \quad (1)$$

Exact expressions for  $\kappa_{||}$  exist for special cases [1, 2]. Otherwise, one calculates numerically, e.g. by integrating the longitudinal electric field in a frame moving with the bunch (Secs.3.2.2, 3.2.3). If the bunch moves with a constant speed  $v < c$ , as e.g. in a proton linac,  $\kappa_{||}$  depends on  $v$ .

The loss factor  $\kappa_{||}$ , usually stated in units  $V/pC = 10^{12} \text{ V/C}$ , is

$$\begin{aligned} \kappa_{||}(\sigma) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} d\omega Z_{||}(\omega) h(\omega, \sigma) \\ &= \frac{1}{\pi} \int_0^{\infty} d\omega \text{Re} Z_{||}(\omega) h(\omega, \sigma) \end{aligned} \quad (2)$$

where  $h(\omega, \sigma) = \tilde{\lambda}(\omega) \tilde{\lambda}^*(\omega)$  is the spectral power density of the bunch of rms length  $\sigma$ .

For a *line* spectrum, the integral becomes an infinite sum. For a single bunch in a circular accelerator,

$$\kappa_{||}(\sigma) = \frac{\omega_0}{2\pi} \sum_{p=-\infty}^{\infty} Z_{||}(p\omega_0) h(p\omega_0, \sigma) \quad (3)$$

For short bunches in large machines ( $\omega_0 \ll 1/\sigma$ ), the sum can be replaced by an integral.

In terms of wake function  $W_{||}(t)$  (Sec.3.2), the loss factor is

$$\kappa_{||} = \int_{-\infty}^{\infty} d\tau W_{||\lambda}(\tau) \lambda(\tau) = \int_0^{\infty} dt W_{||}(t) S(t) \quad (4)$$

where  $W_{||\lambda}(\tau) = \int_0^{\infty} dt W_{||}(t) \lambda(\tau - t)$  is the wake potential,  $S(t) = \int_{-\infty}^{\infty} d\tau \lambda(\tau) \lambda(\tau - t)$  is the auto-correlation function of the line density.

For a Gaussian bunch with rms  $\sigma = \sigma_z/\beta c$ ,  $\lambda = e^{-t^2/2\sigma^2}/(\sqrt{2\pi}\sigma)$ ,  $h(\omega) = e^{-\omega^2\sigma^2}$ ,

$$S(\tau) = \frac{1}{2\sqrt{\pi}\sigma} e^{-\tau^2/4\sigma^2} \quad (5)$$

Examples of auto-correlation functions for other bunch distributions can be found in [3].

In a resonator with shunt resistance  $R_s$  ( $\equiv V_{\text{cav}}^2/(2P_{\text{wall}})$ ), quality factor  $Q_r$ , and resonant frequency  $\omega_r$ ,

$$\begin{aligned} \kappa_{||} &= \frac{R_s}{4Q'_r} [\omega_{1\text{W}}(\omega_1\sigma) - \omega_{2\text{W}}(\omega_2\sigma)] \\ &= \frac{R_s}{2Q'_r} \text{Re} [\omega_{1\text{W}}(\omega_1\sigma)] \quad \text{for } Q_r \geq \frac{1}{2} \end{aligned} \quad (6)$$

### Sec.3.2: IMPEDANCES AND WAKE FUNCTIONS

where  $\omega_{1,2} = (\omega_r/Q_r)[-i/2 \pm Q'_r]$ ,  $|Q'_r|^2 = Q_r^2 - 1/4$ , and  $w(z)$  is the complex error function [4]. The loss factor, normalized by  $\omega_r R_s / 2Q_r$ , is shown in Fig.1a. Approximate expressions for  $\kappa_{\parallel}$  are found below, while the relation between  $\kappa_{\parallel}$  and  $Z_{\parallel}/n$  is discussed in [1].

| Cases                                   | $\kappa_{\parallel}(\sigma) \approx$   |
|---|--|
| high- $Q$ cavity mode ( $Q > 10$ or so) | $\frac{\omega_r R_s}{2Q_r} e^{-\omega_r^2 \sigma^2}$                                     |
| short bunches $\omega_r \sigma \ll 1$   | $\frac{\omega_r R_s}{2Q_r} \left[ 1 - \frac{2}{\pi} \frac{\omega_r \sigma}{Q_r} \right]$ |
| low- $Q$ cavity and long bunches        | $\frac{R_s}{4\sqrt{\pi} Q^2 \omega_r^2 \sigma^3}$  |

**The kick factor** The transverse loss factor for the dipole ( $m = 1$ ) mode, also called the kick factor, with dimensions V/(pC-m), is defined by

$$\begin{aligned} \kappa_{\perp}(\sigma) &= \frac{1}{2\pi i} \int_{-\infty}^{\infty} d\omega Z_{\perp}(\omega) h(\omega, \sigma) \\ &= \frac{1}{\pi} \int_0^{\infty} d\omega \text{Im}Z_{\perp}(\omega) h(\omega, \sigma) \\ &= \int_0^{\infty} dt W_{\perp}(t) S(t) \\ &= \int_{-\infty}^{\infty} d\tau W_{\perp\lambda}(\tau) \lambda(\tau) \end{aligned} \quad (7)$$

in terms of the transverse impedance  $Z_{\perp}(\omega)$ , wake potential  $W_{\perp\lambda}(\tau)$ , and wake function  $W_{\perp}(\tau)$  (Sec.3.2). For a Gaussian bunch traversing a resonator the kick factor, normalized by  $\omega_r R_{\perp} / Q_r$ , is shown in Fig.1b (for  $Q_r \geq 1/2$ ),

$$\begin{aligned} \kappa_{\perp}(\sigma) &= \frac{\omega_r R_{\perp}}{4Q'_r} \left[ w(\omega_1 \sigma) - w(\omega_2 \sigma) \right] \\ &= \frac{\omega_r R_{\perp}}{2Q'_r} \text{Im} \left[ w(\omega_1 \sigma) \right] \end{aligned} \quad (8)$$

For short bunches  $\kappa_{\perp}(\sigma) \approx \frac{\omega_r R_{\perp}}{Q_r} \frac{\omega_r \sigma}{\sqrt{\pi}}$ , while for long ones  $\kappa_{\perp}(\sigma) \approx \frac{R_{\perp}}{Q_r} \frac{1}{2\sqrt{\pi}\sigma}$ .

The parameters of an equivalent transverse broad-band resonator can be estimated from the dependence of the kick factor on bunch length. In particular,  $R_{\perp}/Q_r = 2\sqrt{\pi} \lim_{\sigma \rightarrow \infty} (\omega_r \sigma) \kappa_{\perp}$ , and  $\omega_r = \alpha c / \hat{\sigma}$ , where  $\hat{\sigma}$  is the location of the maximum  $\kappa_{\perp}(\sigma)$ , and  $\alpha$  is a correction factor of order unity [5].

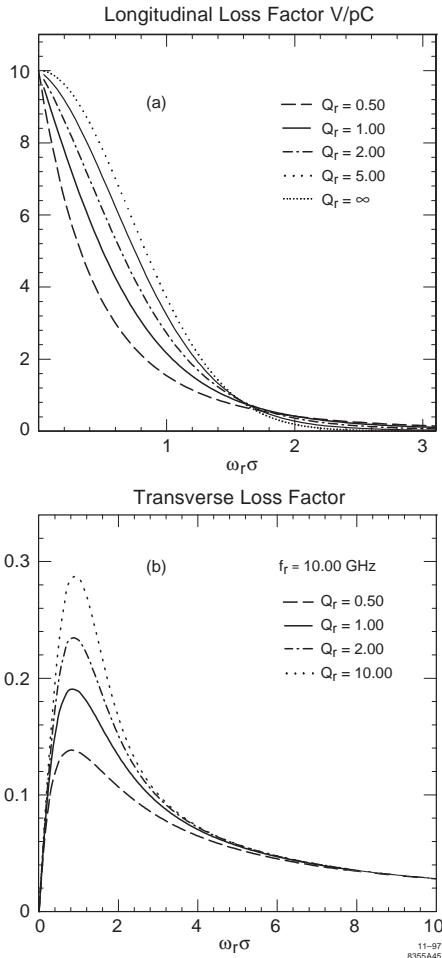


Figure 1: Normalized (a) longitudinal and (b) transverse loss factors of Gaussian bunch in resonator impedances with various quality factors  $Q_r$  versus  $\omega_r \sigma$ .  
11-97  
8355A45

**The modal loss factors** For the  $n$ -th longitudinal mode, the modal loss factor is defined  $k_n = V_n V_n^* / (4U_n)$ , where  $V_n$  is the induced voltage across the impedance and  $U_n$  the stored energy. For a resonator impedance  $Z_{\parallel}$ ,

$$k_n = \frac{1}{\pi} \int_0^{\infty} d\omega \text{Re}Z_{\parallel}(\omega_n) = \frac{\omega_n R_n}{2 Q_n} \quad (9)$$

Behind a point-bunch ( $\tau > 0$ ), the wake function is

$$W_{\parallel}(\tau) = 2 \sum_{n=1}^{\infty} k_n \cos(\omega_n \tau) \quad (10)$$

while it vanishes in front of it ( $\tau < 0$ ). For a Gaussian bunch, the total loss factor is

$$\kappa_{\parallel}(\sigma) = \sum_{n=1}^{\infty} k_n h(\omega_n, \sigma) = \sum_{n=1}^{\infty} k_n e^{-\omega_n^2 \sigma^2} \quad (11)$$

For the  $n$ -th dipole mode, the modal kick factor is  $k_{1n} = \frac{V_{1n} V_{1n}^*}{4U_{1n}}$  where  $V_{1n}$  is the voltage at the iris or beam tube radius. The transverse wake function is

$$W_{\perp}(\tau) = \frac{2c}{b^2} \sum_{m=1}^{\infty} \frac{k_{1n}}{\omega_{1n}} \sin(\omega_{1n}\tau) \quad (12)$$

For a Gaussian bunch, the total kick factor is

$$\begin{aligned} \kappa_{\perp}(\sigma) &= \frac{2c}{b^2} \sum_{m=1}^{\infty} \frac{k_{1n}}{\omega_{1n}} \int_0^{\infty} dt S(t) \sin \omega_{1n} t \\ &= \frac{2c}{\sqrt{\pi} b^2} \sum_{m=1}^{\infty} \frac{k_{1n}}{\omega_{1n}} D(\omega_{1n}\sigma) \end{aligned} \quad (13)$$

where  $D(z) = \exp(-z^2) \int_0^z dt \exp(t^2)$  is the Dawson integral [4].

**Examples other than resonator impedance**  
For further examples (other vacuum chamber discontinuities, such as tapers and steps) [2, 6].

Resistive wall For a cylindrical resistive wall with conductivity  $\sigma_c$  and pipe radius  $b$ , the loss factor per unit length for a Gaussian bunch is

$$\frac{\kappa_{\parallel}(\sigma)}{L} = \frac{\Gamma(\frac{3}{4})c}{4\pi^2 b \sigma_c^{3/2}} \left( \frac{Z_0}{2\sigma_c} \right)^{1/2} \quad (14)$$

$\Gamma(3/4) \approx 1.225$ . Eq.(14) holds also for a beam moving (a) centered between two parallel plates spaced by  $2b$ , and (b) a distance  $b$  from a single resistive plate. For an elliptical pipe,  $\kappa_{\parallel}$  is minimum when (major axis)/(minor axis)  $\approx 1.4$ , and the minimum value is 0.87 of Eq.(14) [7].

Collimator For a collimator of aperture  $a$  in a pipe of radius  $b$ , and  $c\sigma \ll a$ ,  $\kappa_{\parallel} \approx Z_0 \ln(b/a)/(2\pi^{3/2}\sigma)$ .

Very short bunches For very short bunches in a linac structure,  $\kappa_{\parallel}$  is not determined by the collective interaction of the beam with the periodic cavity array, but by diffractive losses in the individual cells [6]. In this regime ( $\sigma < a^2/(Np)$ , with  $a$  iris aperture radius,  $p$  periodic length,  $N$  number of cells in the structure),  $\kappa_{\parallel}$  diverges as  $\sigma^{-1/2}$  (see Secs.3.2.2, 3.2.4).

For a resonant cavity in a storage ring,  $\kappa_{\parallel}$  is given by a sum over modes up to the cut-off frequency, plus a high frequency diffraction contribution (diffraction model [8], also Sec.3.2.4)

$$\kappa_{\parallel} \approx \frac{\Gamma(\frac{1}{4})Z_0}{4\pi^{5/2}b} \sqrt{\frac{cg}{\sigma}} \quad (15)$$

where  $b$  is the vacuum chamber radius,  $g$  is the axial length of the cavity.

For perfectly conducting walls,  $\text{Re } Z_{\parallel}$  is zero below the first cavity resonance at  $\omega_{01}/c = j_{01}/d$ , and  $\kappa_{\parallel}$  decreases exponentially with  $\sigma$ , rather than  $\sigma^{-3}$  in the resonator model.

For a cavity-like structure with a small opening gap (smaller than the chamber radius) and for a short bunch (shorter than the chamber radius), the loss factor can be estimated in a good accuracy only from the gap size, the chamber radius and the bunch length, regardless of the exact shape and size of the structure. This is because the bunch does not have enough time to see the details of the structure in one passage, and the interaction takes place almost solely between the gap and the bunch. The analytical expression of the loss factor in this case is available in a form suitable for numerical evaluations [9].

Structures without cylindrical symmetry For vacuum chambers not cylindrically symmetric but with mirror symmetry about two orthogonal planes, intersecting at an axis, the loss factors depend on beam displacement. The kick factors of off-axis beams have finite cross terms and can become negative (focusing).

For structures with less or no symmetry, no axis is defined, and the loss factors must be computed along straight trajectories, e.g. parallel to the vacuum chamber wall.

## References

- [1] A.W. Chao, Physics of Collective Beam Instabilities in High Energy Accelerators, Wiley (1993)
- [2] B. Zotter, S. Kheifets, Impedances and Wakes in High Energy Particle Accelerators, World Scientific (1997)
- [3] R. Bracewell, The Fourier Transform and Its Applications, McGraw-Hill (1965) 152
- [4] M. Abramowitz, I. Stegun, Handbook of Mathematical Functions (1965)
- [5] B. Zotter, CERN Report LEP-TH/87-34 (1987)
- [6] S.A. Heifets, S.A. Kheifets, PR D39 (1989) 960
- [7] A. Piwinski, DESY 94-068 (1994)
- [8] K. Bane, M. Sands, SLAC-PUB-4441 (1987)
- [9] Y. Shobuda, Y.H. Chin, K. Takata, PRST-AB 10, 044403 (2007)

### 3.2.7 Trapped Modes *S.S. Kurennoy, LANL*

The term “trapped mode” refers to a mode which cannot propagate in the beam pipe. Often this

### Sec.3.3: PARTICLE-MATTER INTERACTION

term is reserved only for modes whose frequency is above the pipe cutoff frequency [1], since modes below the pipe cutoff cannot propagate in the pipe anyway. An example of a trapped mode is a mode in a long cavity, e.g., multi-cell cavity or IR, with fields concentrated near the cavity center and vanishing toward its ends. Such a mode does not propagate into the pipe due to its very weak coupling to the pipe modes. The trapped modes usually produce narrow resonance peaks of the coupling impedance. In normal-conducting cavities such resonances can be damped (coupled out), but suppressing them in superconducting cavities is complicated. A detailed analysis of the trapped modes in multi-cell cavities, supported both by numerical and experimental studies, can be found in [2, 3].

In other cases the term “trapped mode” is generalized by including the modes with frequencies slightly below the cutoff [4]. A small discontinuity (e.g. a pillbox or a hole, but not an iris or a post) on a smooth pipe can produce such a mode localized near the discontinuity [4]. For a small axisymmetric bulge with the area  $A$  in its longitudinal cross section, the frequency shift  $\Delta\omega_{01}$  down from the cutoff frequency  $\omega_{01} = \mu_{01}c/b$  of the  $TM_{01}$  mode in the cylindrical pipe of radius  $b$  is

$$\Delta\omega_{01} = \omega_{01} \frac{\mu_{01}^2}{2} \left( \frac{A}{b^2} \right)^2 \quad (1)$$

where  $\mu_{01}$  is the first root of the Bessel function  $J_0(x)$ . The damping rate  $\gamma_{01}$  for the case of a finite, though large, conductivity  $\sigma$  of the walls, is  $\gamma_{01} = \omega_{01}\delta_s/(2b)$ , where  $\delta_s$  is the skin depth. The trapped mode disappears when  $\gamma_{01}$  becomes larger than  $\Delta\omega_{01}$ , but in the limit of perfectly conducting walls it exists even for a very small perturbation. This trapped mode leads to a resonance of the longitudinal impedance with the shunt impedance

$$R_{01} = \frac{Z_0\mu_{01}^3 A^3}{\pi\delta_s b^5} \quad (2)$$

The results for an enlargement remain valid for a hole if one replaces  $A \rightarrow \alpha_\theta/(4\pi b)$  in Eqs.(1) and (2), where  $\alpha_\theta$  is the magnetic susceptibility of the hole (Sec.3.2.4). The resonances due to the trapped modes can be dangerous in large superconducting colliders like the LHC, where the vacuum chamber incorporates the inner thermal screen (liner) with many pumping holes. The trapped-mode effect for many holes has been studied analytically [5], and was confirmed by

measurements of the trapped modes in perforated pipes [6]. The trapped modes produced by small discontinuities in vacuum chambers with arbitrary homogeneous cross sections have been investigated in [7].

## References

- [1] S. Heifets, S. Kheifets, RMP 63 (1991) 631
- [2] T. Weiland et al, PRST-AB 2, 042001 (1999)
- [3] R. Schuhmann, T. Weiland, PRST-AB 3, 122002 (2000)
- [4] G.V. Stupakov, S.S. Kurennoy, PR E49 (1994) 794
- [5] S.S. Kurennoy, PR E51 (1995) 2498
- [6] F. Caspers, T. Scholz, PA 51 (1995) 251
- [7] S.S. Kurennoy, R.L. Gluckstern, G.V. Stupakov, PR E52 (1995) 4354

### 3.3 PARTICLE-MATTER INTERACTION

In this section, effects caused by beam-beam collisions and by beam interactions with residual gas and accelerator components are considered. These effects lead to decrease of beam intensity, luminosity deterioration in colliders, and formation of beam halo. Loss of beam particles further leads to (i) beam induced gas desorption; (ii) irradiation of conventional and superconducting components with a possible quench; (iii) increase of background rates in the detectors; (iv) possible radiation damage, overheating or destruction of components; (v) increased prompt and residual radiation fields in the machine and environment.

#### 3.3.1 Basic Formulae [1]

*M. Tigner, Cornell U.  
A.W. Chao, SLAC*

Symbols  $M$  incident particle mass,  $E = \gamma Mc^2$  incident particle energy,  $T$  kinetic energy,  $ze$  charge of incident particle,  $Z, A$  atomic number and atomic mass of medium,  $I$  mean excitation energy,  $\delta$  density effect correction to ionization energy loss,  $\hbar\omega_p = \sqrt{4\pi N_e r_e^3 m_e c^2}/\alpha = 28.816\sqrt{\rho Z/A}$  [eV] plasma energy ( $\rho$  in g/cm<sup>3</sup>),  $N_e$  electron volume density (m<sup>-3</sup>),  $K/A = 4\pi N_A r_e^2 m_e c^2/A$ ,  $X_0$  radiation length in g/cm<sup>2</sup>,  $E_c$  critical energy in MeV.

**Single Coulomb scattering of spin- $\frac{1}{2}$  particles**  
Neglecting atomic electrons and nuclear form factor effects and assuming scattering nucleus at rest

in lab [2],

$$\frac{d\sigma}{d\Omega} = \frac{Z^2}{4} r_e^2 \left( \frac{m_e c}{\beta p} \right)^2 \frac{1}{\sin^4 \frac{\theta}{2}} \left( 1 - \beta^2 \sin^2 \frac{\theta}{2} \right) \quad (1)$$

When screening is included, and for small angles, Eq.(1) reads

$$\frac{d\sigma}{d\Omega} \approx 4 Z^2 r_e^2 \left( \frac{m_e c}{\beta p} \right)^2 \frac{1}{(\theta^2 + \theta_{\min}^2)^2} \quad (2)$$

where  $\theta_{\min} \approx \alpha Z^{1/3} (m_e c / p)$ .

**Thomson cross section** Unpolarized EM radiation on free electron or proton [3],

$$\frac{d\sigma}{d\Omega} = \frac{r_e^2}{2} (1 + \cos^2 \theta) \quad (3)$$

$$\sigma_T = \frac{8\pi}{3} r_e^2 \quad (4)$$

**Klein-Nishina cross section** For high energies  $\hbar\omega_{x-ray} \gg m_e c^2$  [3],

$$\frac{\sigma_{KN}}{\sigma_T} \approx \frac{3}{4} \frac{mc^2}{\hbar\omega} \left[ \frac{1}{4} + \frac{1}{2} \ln \left( \frac{2\hbar\omega}{m_e c^2} \right) \right] \quad (5)$$

See also Sec.7.6.1.

**Ionization energy loss** Bethe-Bloch equation for mean rate of energy loss (stopping power) for moderately relativistic charged particles other than electrons,

$$-\frac{dE}{dx} = K z^2 \frac{Z}{A} \frac{1}{\beta^2} \quad (6)$$

$$\times \left[ \frac{1}{2} \ln \frac{2m_e c^2 \beta^2 \gamma^2 T_{\max}}{I^2} - \beta^2 - \frac{\delta}{2} \right]$$

Here

$$K/A = 4\pi N_A r_e^2 m_e c^2 / A$$

$$= 0.307 \text{ MeV cm}^2 \text{g}^{-1} \text{ for } A = 1 \text{ g/mol}$$

$$T_{\max} = \frac{2m_e c^2 \beta^2 \gamma^2}{1 + 2\gamma m_e / M + (m_e / M)^2}$$

is the maximum kinetic energy which can be imparted to a free electron in a single collision. The units are so that  $dx$  is measured in mass per unit area ( $\text{g cm}^{-2}$ ). The mean excitation energy  $I$  is  $(10 \pm 1 \text{ eV}) \times Z$  for elements heavier than oxygen. At very high energies, the density effect correction  $\delta/2 \rightarrow \ln(\hbar\omega_p/I) + \ln \beta\gamma - 1/2$ . For pions on various material, see Fig.1. Eq.(6) may be integrated to find the total range  $R$ . See Fig.2.

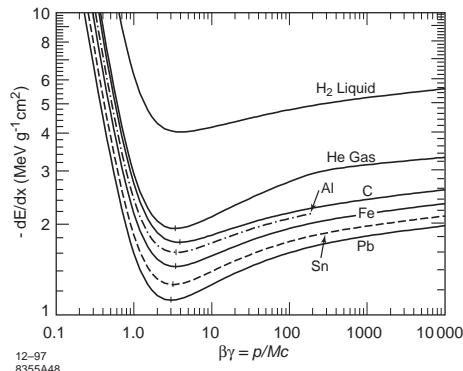


Figure 1: Energy loss rate in liquid hydrogen, gaseous helium, carbon, aluminum, tin, and lead.

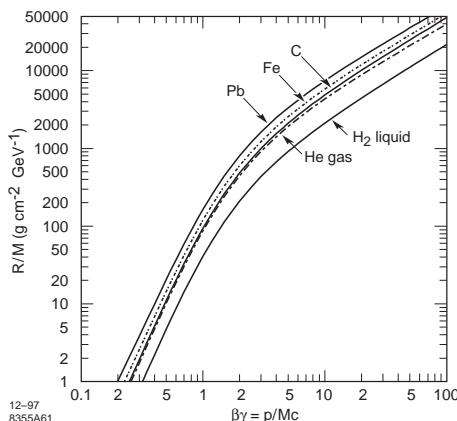


Figure 2: Range of heavy charged particles in liquid (bubble chamber) hydrogen, helium gas, carbon, iron, and lead.

**Multiple Coulomb scattering through small angles** If we define  $\theta_0 = \theta_{\text{plane}}^{\text{rms}} = \frac{1}{\sqrt{2}} \theta_{\text{space}}^{\text{rms}}$ , then it is often sufficient to use a Gaussian approximation for the central 98% of the projected angular distribution, with a width

$$\theta_0 = \frac{13.6 \text{ MeV}}{\beta c p} z \sqrt{\frac{x}{X_0}} \left( 1 + 0.038 \ln \frac{x}{X_0} \right) \quad (7)$$

Here  $p$ ,  $\beta c$ , and  $z$  refer to the incident particle;  $x$  is the thickness of the scattering medium. This value of  $\theta_0$  is from a fit to Molière distribution for singly charged particles with  $\beta = 1$  for all  $Z$ , and is accurate to 11% or better for  $10^{-3} < (x/X_0) < 100$ .

### Sec.3.3: PARTICLE-MATTER INTERACTION

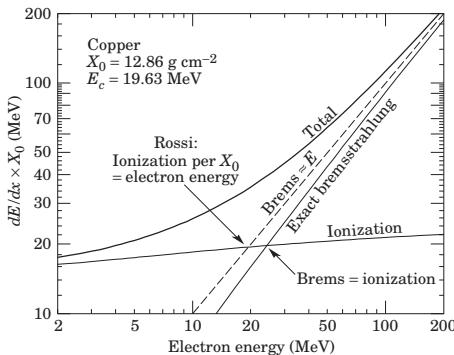


Figure 3: Two definitions of the critical energy  $E_c$ .

**Radiation length** This is the mean distance over which a high-energy electron loses all but  $1/e$  of its energy by bremsstrahlung, and is the scale length for high-energy electromagnetic cascades (other than hydrogen),

$$\frac{1}{X_0} = 4\alpha r_e^2 \frac{N_A}{A} \left\{ Z^2 [L_{\text{rad}} - f(Z)] + Z L'_{\text{rad}} \right\} \quad (8)$$

$$\begin{aligned} f(Z) &\approx a^2 [(1+a^2)^{-1} + 0.20206 \\ &\quad - 0.0369 a^2 + 0.0083 a^4 - 0.002 a^6] \\ &\quad (a = \alpha Z) \end{aligned} \quad (9)$$

$$L_{\text{rad}} = \ln \frac{184.15}{Z^{1/3}}, \quad L'_{\text{rad}} = \ln \frac{1194}{Z^{3/2}} \quad (10)$$

For  $A = 1 \text{ g/mol}$ ,  $4\alpha r_e^2 N_A / A = (716.4 \text{ g/cm}^2)^{-1}$ .

An electron loses energy by bremsstrahlung at a rate nearly proportional to its energy, while the ionization loss rate varies only logarithmically with the electron energy. The *critical energy*  $E_c$  is sometimes defined as the energy at which the two loss rates are equal,  $E_c = (800 \text{ MeV})/(Z + 1.2)$ . Among alternate definitions of  $E_c$  is that of Ref.[4], which defines  $E_c$  as the energy at which the ionization loss per radiation length is equal to the electron energy. Equivalently, it is the same as the first definition with the approximation  $|dE/dx|_{\text{brems}} \approx E/X_0$ . These definitions are illustrated in the case of copper in Fig.3.

**Muon energy loss at high energy** At sufficiently high energies, radiative processes become more important than ionization for all charged particles. For muons and pions in materials such as iron, this “critical energy” occurs at several hundred GeV. See Fig.4.

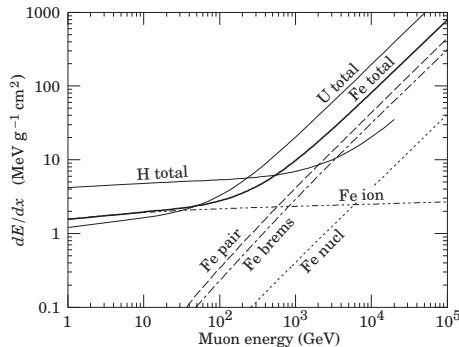


Figure 4: Average energy loss of a muon in hydrogen, iron, and uranium as a function of muon energy.

### Bremsstrahlung on gas nuclei [2]

$$\begin{aligned} \frac{d\sigma}{du} &\approx 4\alpha r_e^2 Z(Z+1) \\ &\times \frac{4}{3u} (1-u+.75u^2) \ln \left( \frac{184.15}{Z^{1/3}} \right) \end{aligned}$$

with  $u = k/E$  the fractional energy loss by radiation, where  $k$  is the energy of the photon radiated in the bremsstrahlung event. Integral forms are found in Secs.3.1.12, 3.3.2 and 4.6.

### Beam-beam bremsstrahlung loss cross section [2] (See also Sec.2.5.3.2.)

$$\begin{aligned} \sigma_{\text{bbb loss}} &= \int_{k_{\min}}^E \frac{d\sigma}{dk} dk \\ &\approx \frac{16\alpha r_e^2}{3} \left[ \left( \ln \frac{E_{\text{cm}}^2}{m_e^2 c^4} - \frac{1}{2} \right) \left( \ln \frac{E_{\text{beam}}}{k_{\min}} - \frac{5}{8} \right) \right. \\ &\quad \left. + \frac{1}{2} \left( \ln \frac{E_{\text{beam}}}{k_{\min}} \right)^2 - \frac{3}{8} - \frac{\pi^2}{6} \right] \end{aligned} \quad (11)$$

where  $k_{\min}$  is the largest energy change that can be tolerated by the acceptance of the accelerator.

### References

- [1] K. Nakamura et al (Particle Data Group), J. Phys. G37, 075021 (2010)
- [2] F.C. Porter, NIM A302 (1991) 209
- [3] J.D. Jackson, Classical Electrodynamics, 3rd ed., Wiley (1999)
- [4] B. Rossi, High Energy Particles, Prentice-Hall (1952)

### 3.3.2 Beam and Luminosity Lifetime

#### 3.3.2.1 Protons

N.V. Mokhov, V.I. Balbekov, FNAL

One needs to distinguish: *Beam lifetime*, *Emitance lifetime*, and (in colliders) *Luminosity lifetime*. Involving phenomenon is *beam halo* formation that depends on partial effects  $i$ . The total lifetime  $\tau_{total}$  is calculated from

$$\frac{1}{\tau_{total}} = \sum_i \frac{1}{\tau_i} \quad (1)$$

Partial effects exclusively due to beam interactions with residual gas and particle collisions at the interaction point (IP) are considered in this section. Other related effects are considered elsewhere: *Touschek effect* (Sec.2.4.12) and *intra-beam scattering* (Sec.2.4.12); *Beam cooling* (Secs.2.7,4.16); *Beam-beam scattering in colliders* (Sec.4.7.6.2); and other *beam-beam effects* (Sec.2.5.1,4.13,4.14).

**Beam lifetime** is determined by single-acting processes: beam-gas interactions and (in colliders) by beam-beam collisions at the IP,  $1/\tau_{beam} = 1/\tau_{gas} + 1/\tau_{IP}$ . As a result of beam-gas single electromagnetic and nuclear elastic scatterings with angles exceeding the angular beam size, and nuclear inelastic interactions, particles leave the core creating beam halo and ultimately are lost on the limiting apertures of the machine. Diffusion caused by multiple Coulomb scattering (MCS) on electrons and nuclei of residual gas atoms is negligible at  $E > \text{a few GeV}$  [1] (Sec.4.7.6.2). The resulting “beam profile” can be drastically different from Gaussian. Fig.1 shows a transverse profile measured at the 70 GeV proton synchrotron at IHEP (Protvino) and fitted for different residual gas pressures  $P$  by [2]

$$N(x) = e^{-x^2/2\sigma^2} + C_0 \ln(P/P_0) \frac{P}{x^2} \quad (2)$$

where  $x$  is a distance from the beam axis in mm,  $(3 < x < 40 \text{ mm})$ ,  $\sigma = 2.2 \text{ mm}$ ,  $C_0 = -4.8 \times 10^4 \text{ mm}^2/\text{Torr}$ ,  $P < 10^{-3} \text{ Torr}$ , and  $P_0 = 10^{-3} \text{ Torr}$ .

At high energies elastic and diffractive scatterings contribute to the beam halo and emittance growth, while inelastic nuclear interactions result in multi-particle production and shower development along one or two magnets. Residual gas is

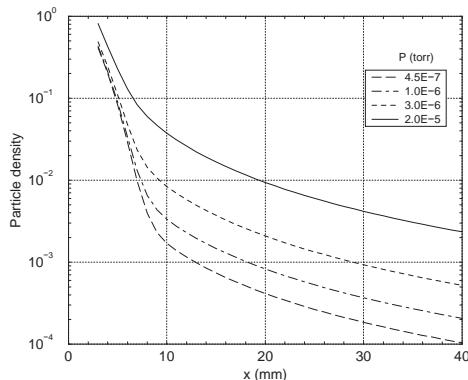


Figure 1: Transverse beam profile in 70 GeV proton synchrotron for several values of residual gas pressure.

usually a mixture of H, C, N and O atoms. At high energies, the differential cross-section of proton-hydrogen (antiproton-hydrogen) elastic scattering is well described by a simple exponent

$$d\sigma/dt \sim e^{-b|t|} \quad (3)$$

where  $t \approx -p^2\theta^2$  is the 4-momentum transfer squared,  $p$  is the proton momentum, in units of  $(\text{GeV}/c)^2$ . At  $0.001 < |t| < 0.8 (\text{GeV}/c)^2$ ,  $b$  increases with energy from  $12 (\text{GeV}/c)^{-2}$  at  $\sqrt{s} = 10 \text{ GeV}$  up to  $16.4 (\text{GeV}/c)^{-2}$  at  $\sqrt{s} = 1.8 \text{ TeV}$ , where  $s$  is the CM collision energy squared [3].

The proton-nucleus elastic differential cross-section on light and medium nuclei ( $9 < A < 62$ ) at  $p > 5 \text{ GeV}/c$  is described as [4]

$$\frac{d\sigma}{d\Omega} [\text{mb/str}] = 38.0 p^2 A^{1.63} e^{-14.5 A^{0.66}|t|} + 53.2 p^2 A^{0.33} e^{-10|t|} \quad (4)$$

Corresponding rms angle  $\theta_{el} = C/p$ , where

$$C = \sqrt{\frac{0.0595 A^{-0.02} + 0.175}{0.862 A^{0.64} + 1.75}}$$

For nitrogen  $C = 0.19 \text{ GeV}/c$ . Using Eq.(4), one finds that, e.g. at Tevatron ( $p = 980 \text{ GeV}/c$ ), most of elastic (and certainly all inelastic) scatterings force particles to leave the core creating halo and beam loss, with

$$\theta_{el} \sim 0.2 \text{ mrad} \gg \theta_{beam} = 0.011 \text{ mrad}$$

where  $\theta_{beam}$  is rms beam divergence.

Summing over species of circumferentially averaged gas densities  $n_i$ , one gets

$$\frac{1}{\tau_{gas}} = -\frac{1}{N} \frac{dN}{dt} = v \sum_i \sigma_i n_i \quad (5)$$

### Sec.3.3: PARTICLE-MATTER INTERACTION

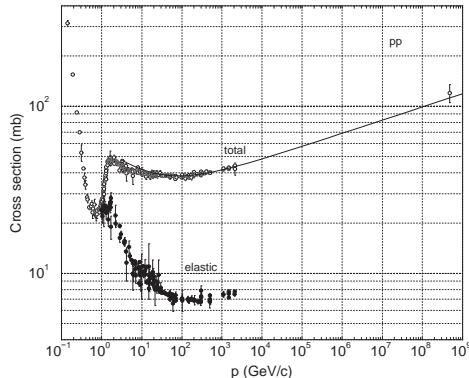


Figure 2: Total and elastic pp cross-sections [5].

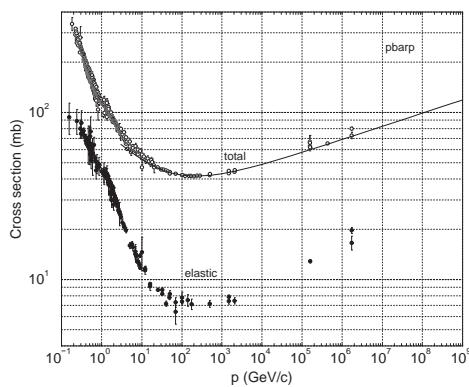


Figure 3: Total and elastic p $\bar{p}$  cross-sections [5].

where  $v = \beta c$  is the proton velocity, and  $\sigma_i$  is the total proton or antiproton cross-section on species  $i$ . Experimental data on the total and elastic cross-sections for pp or p $\bar{p}$  collisions are presented in Figs.2-3 along with the Regge description [5]

$$\sigma_{\text{tot}}[\text{mb}] = 22 s^{0.079} + Y s^{-0.46} \quad (6)$$

where  $Y_{\text{pp}} = 56.1$ ,  $Y_{\text{p}\bar{p}} = 98.2$ . Fig.4 shows total proton-nucleus cross-section calculated in [6] within a multichannel generalization of the Glauber model using  $\sigma_{\text{pp}}$  parameterizations of experimental data and the Woods-Saxon distribution of nucleons in nuclei. The total p- $A$  cross-section remains approximately constant at proton momenta  $1 \text{ GeV}/c < p < 50 \text{ TeV}/c$ .

For nitrogen  $\sigma \approx 410 \text{ mb}$  at proton momenta  $3 \text{ GeV}/c < p < 5 \text{ TeV}/c$ . Summing over the beam

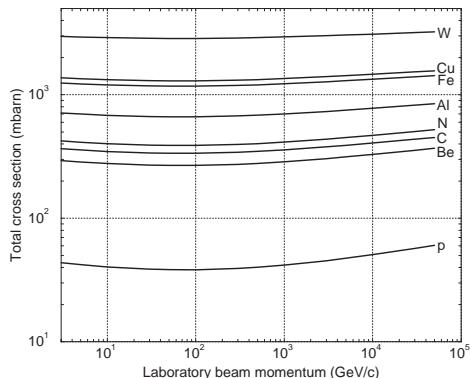


Figure 4: Total proton-nucleus cross-sections [6].

pipe regions of “constant” pressure and temperatures, Eq.(5) for molecular nitrogen becomes

$$\frac{1}{\tau_{\text{gas}}[\text{s}]} \approx 2.4 \times 10^5 \beta \sum_k \frac{w_k P_k}{T_k} \quad (7)$$

where  $P_k$  (Torr) is the gas pressure in the pipe of the  $k$ th region of the fraction  $w_k$  of the total circumference at the temperature  $T_k$  (K), relative velocity  $\beta \approx 1$ . For a gas mixture one should use the equivalent nitrogen pressure

$$P_{\text{eq}}^{\text{Nucl}} = 0.0861 \sum_i P_i \sum_j k_{ij} A_{ij}^{2/3} \quad (8)$$

where  $k_{ij}$  is the number of  $j$  atoms in  $i$  molecule and  $A_{ij}$  is atomic mass of  $j$  atom in  $i$  molecule.

In colliders, all inelastic beam-beam interactions remove particles from each beam. Only a fraction of elastic events at IP has the same effect. For example, using Eq.(3) for the IP of Tevatron,

$$\theta_{\text{el}} \sim 0.17 \text{ mrad}, \quad \theta_{\text{beam}} \sim 0.09 \text{ mrad} \quad (9)$$

and about 40% of elastically scattered protons (anti-protons) remain in the  $3\sigma$  core after a bunch-bunch collision at the IP. So, the beam lifetime due to collisions is given by a sum over interaction rates at each of the  $n$  IP in the collider,

$$\frac{1}{\tau_{\text{IP}}} = \frac{1}{n_B N_B} \sum_n \sigma_{\text{pp}} \mathcal{L}_n \quad (10)$$

where  $n_B$  is the number of bunches,  $N_B$  is the number of particles per bunch,  $\sigma_{\text{pp}}$  is the inelastic plus a fraction of elastic (60% in the Tevatron case, 9 mb) pp or p $\bar{p}$  cross-section, and  $\mathcal{L}_n$  is the average luminosity at the  $n$ th IP.

**Emittance lifetime** in colliders, assuming identical beams, is defined as  $1/\tau_\epsilon = (1/\epsilon)d\epsilon/dt$ , with  $\epsilon$  being the rms normalized beam emittance. Under normal conditions the growth of  $\epsilon$  is the dominant cause of luminosity deterioration [7]. In an “ideal” machine, the growth of emittance is due to beam-gas MCS, beam-beam elastic scattering at the IPs, and the intrabeam scattering. The beam-gas emittance growth over a given time period is proportional to the residual gas pressure and the average  $\beta$ -function [8]. Summing again over the beam tube regions of “constant” pressure and temperatures, the emittance growth rate due to MCS, e.g. in  $x$ -plane, is (see also Sec.4.7.6)

$$\frac{1}{\epsilon_x} \frac{d\epsilon_x}{dt} = \frac{\pi}{2\epsilon_x} \beta \gamma \sum_k w_k \beta_{xk} \theta_{tk}^2 \quad (11)$$

where  $\epsilon_x$  is the normalized rms beam emittance (phase space area with a factor of  $\pi$  included). Here  $\theta_{tk}^2$  is the rms angle of MCS per time unit in the pipe of the  $k$ th region, and  $\beta_{xk}$  is average  $\beta$ -function in  $x$ -plane in the  $k$ th region. Summing over species of average gas densities  $n_i$  in each region, one gets

$$\theta_t^2 = \frac{4\pi r_p^2 c}{\beta^3 \gamma^2} \sum_i n_i Z_i (Z_i + 1) \ln(183 Z_i^{-1/3}) \quad (12)$$

where  $Z_i$  is atomic number of species, and  $r_p = 1.54 \times 10^{-16}$  cm. For molecular nitrogen at pressure  $P$  (Torr) and temperature  $T$  (K),

$$\frac{1}{\epsilon_x} \frac{d\epsilon_x}{dt} \approx \frac{2.2 \times 10^7}{\epsilon_x \beta^2 \gamma} \sum_k \frac{w_k \beta_{xk} P_k}{T_k} \quad (13)$$

where time is in seconds,  $\beta_{xk}$  in meters and  $\epsilon_x$  in  $\mu$ m-mrad.

For arbitrary mixture of  $i$  gases with partial pressures  $P_i$ , one uses the nitrogen-equivalent pressure referred to temperature of 273 K in each machine region

$$P_{\text{eq}}^{\text{Coul}} \approx 2 \times 10^{-3} \sum_i P_i \quad (14)$$

$$\times \sum_j k_{ij} Z_{ij} (Z_{ij} + 1) \ln(183 Z_{ij}^{-1/3})$$

In a collider, elastic scattering off the oncoming beam particles at the IP also causes emittance growth. (Intra-beam scattering is discussed in Sec.2.4.12.) Eq.(11) can be applied replacing  $\beta$ -functions averaged over machine regions with  $\beta^*$  at each IP and using the following expression for the rms scattering angle in one of the beams

$$\langle \theta_t^2 \rangle = \sigma'_{\text{el}} \theta_{\text{el}}^2 \frac{\mathcal{L}}{n_B N_B} \quad (15)$$

where  $\beta^*$  and  $\mathcal{L}$  are the  $\beta$ -function and luminosity at the IP,  $\sigma'_{\text{el}}$  is the fraction of the elastic scattering cross-section resulting in emittance growth, and  $\theta_{\text{el}}$  is the rms projected elastic scattering angle, obtained from Eq.(3). This gives (see also Eq.(14), Sec.4.7.6)

$$\frac{1}{\epsilon_N} \frac{d\epsilon_N}{dt} = \frac{1}{8} \frac{f_0 N_B \gamma}{\epsilon_N^2} \sigma'_{\text{el}} \theta_{\text{el}}^2 \quad (16)$$

where  $f_0$  is the revolution frequency.

**Luminosity lifetime** in colliders is a sum of (7), (10), (11) and (16),

$$\frac{1}{\tau_{\mathcal{L}}} = \frac{1}{\tau_{\text{beam}_1}} + \frac{1}{\tau_{\text{beam}_2}} + \frac{1}{\tau_\epsilon} \quad (17)$$

As stated at the beginning, the role of partial contributions is different depending on the machine. For example, for the Tevatron collider, beam-gas interactions and  $p\bar{p}$  collisions at the IPs are the dominant sources in this equation for the processes considered in this section. In reality, processes listed at the beginning as well as the abort kicker system performance, the low- $\beta$  quadrupole squeeze process, Main Ring Lambertson power supply noise, the main dipole/quadrupole power supply ripple and the ambient ground motion [9] all could result in reduction of the luminosity lifetime. Fortunately, many of these sources can be (and in the Tevatron have been) mitigated through studies and improvements.

## References

- [1] A.N. Kalinovskii, N.V. Mokhov, Yu.P. Nikitin, Passage of High-Energy Particles through Matter, AIP (1989)
- [2] A.A. Asseev et al, IHEP 79-91, Protvino (1979)
- [3] R. Rubinstein, Fermilab-Conf-90/160-E (1990)
- [4] J. Ranft, PA 3 (1972) 129
- [5] Review of Particle Properties, PR D54 (1996)
- [6] S.I. Striganov, IHEP 94-14, Protvino (1994)
- [7] G. Jackson et al, Fermilab-Conf-88/80 (1988)
- [8] C. Bovet et al, CERN/MPS-SI/Int. DL/70/4 (1970)
- [9] G. Jackson, AIP Proc. 255 (1992) 105

### 3.3.2.2 Electrons

M.S. Zisman, LBNL

**Symmetric collider** If the beam sizes remain constant as the luminosity decreases (not true when beam-beam limit is reached), the time dependence of  $\mathcal{L}(t)$  is contained entirely in the time

### Sec.3.3: PARTICLE-MATTER INTERACTION

dependence of  $N(t)$ . (See also Sec.4.1.) There are then two distinct processes for luminosity loss: (i) those due to collisions between the two beams, and (ii) single-beam loss processes. The processes in (i) relevant to a high-energy  $e^+e^-$  collider are Bhabha scattering (BS) and radiative Bhabha scattering (RBS) (Sec.3.3.3). In BS process, a beam particle is lost if its angular deflection is beyond the ring's transverse acceptance; in RBS process, loss occurs if the beam particle's momentum change is outside the longitudinal acceptance of the ring.

For a symmetric collider (two beams of identical sizes and intensity) with a loss cross section  $\sigma$  (Sec.3.3.1), the particle loss rate is

$$\frac{dN}{dt}(t) = -\sigma \mathcal{L}(t) = -\sigma k N^2(t) \quad (18)$$

where  $k = \mathcal{L}_0/N_0^2 = f_c/(4\pi\sigma_{x,y}^*\sigma_y^*)$  with  $f_c$  the collision frequency,  $\sigma_{x,y}^*$  the beam sizes at the IP, and subscript 0 means initial values at  $t = 0$ . Solving Eq.(18) yields [1]

$$\begin{aligned} N(t) &= N_0 \left[ \frac{1}{1 + (\sigma \mathcal{L}_0/N_0)t} \right] \\ \mathcal{L}(t) &= \mathcal{L}_0 \left[ \frac{1}{1 + (\sigma \mathcal{L}_0/N_0)t} \right]^2 \end{aligned} \quad (19)$$

The time for the luminosity to decay to a fraction  $f$  of its initial value is

$$t_f = \frac{N_0}{\sigma \mathcal{L}_0} \left( \frac{1}{\sqrt{f}} - 1 \right) \quad (20)$$

**Asymmetric collider** Generalizing to a two-ring asymmetric collider (still assuming constant beam sizes), the particle loss is described by [1]

$$\begin{aligned} \frac{dN_+}{dt}(t) &= -\sigma_+ k N_+ N_- \\ \frac{dN_-}{dt}(t) &= -\sigma_- k N_+ N_- \end{aligned} \quad (21)$$

$$k = \frac{\mathcal{L}_0}{N_0^+ N_0^-} = \frac{f_c/2\pi}{\sqrt{(\sigma_{x+}^* + \sigma_{x-}^*)^2 (\sigma_{y+}^* + \sigma_{y-}^*)}}$$

where we no longer assume that the loss cross sections for the two beams are identical. The solution to Eq.(21) is [1]

$$\begin{aligned} N_+(t) &= N_{0+} \left( \frac{1-r}{e^{Gt}-r} \right) \\ N_-(t) &= N_{0-} \left( \frac{1-r}{1-re^{-Gt}} \right) \\ \mathcal{L}(t) &= \mathcal{L}_0 e^{Gt} \left( \frac{1-r}{e^{Gt}-r} \right)^2 \end{aligned} \quad (22)$$

where

$$G = \mathcal{L}_0 \left( \frac{\sigma_+}{N_{0+}} - \frac{\sigma_-}{N_{0-}} \right), \quad r = \frac{N_{0+}\sigma_-}{N_{0-}\sigma_+} \quad (23)$$

The time for the luminosity to decay to a fraction  $f$  of its initial value is

$$\begin{aligned} t_f &= \frac{1}{G} \ln \left\{ \frac{1}{2f} \left[ (1-r)^2 + 2fr \right. \right. \\ &\quad \left. \left. + (1-r)\sqrt{(1-r)^2 + 4fr} \right] \right\} \end{aligned} \quad (24)$$

**Other loss processes** A full treatment of luminosity lifetime must also take into account the single-beam losses that can independently decrease  $N_+$  and  $N_-$ . These include gas scattering (both elastic and inelastic i.e. gas bremsstrahlung), Touschek scattering (Sec.2.4.12), and, at high energy, Compton scattering off thermal photons (Sec.3.3.4) and beamstrahlung (Sec.2.5.3). In a simple model [2], the gas scattering rate is independent of beam intensity, leading to exponential decay,

$$\frac{1}{\tau_i} = -\frac{1}{N} \frac{dN}{dt} = n_g \beta c \sigma_i \quad (25)$$

where  $n_g = 9.656 \times 10^{24} n_Z P [\text{Torr}] / T [\text{K}]$  is the number density of residual gas atoms,  $\beta c$  is the beam velocity,  $n_Z$  is the number of atoms per gas molecule, and  $\sigma_i$  is the cross section, where  $i$  denotes either the elastic scattering (el) or bremsstrahlung (Br) process. The elastic cross section is [3]

$$\sigma_{\text{el}} = \frac{2\pi r_e^2 Z^2 \beta_\perp}{\gamma^2 A_\perp} \quad (26)$$

where  $A_\perp$  and  $\beta_\perp$  are the ring limiting acceptance (beam chamber radius squared divided by  $\beta_\perp$ ) and corresponding transverse  $\beta$ -function, respectively. The bremsstrahlung cross section in terms of the radiation length  $X_0$  (Sec.3.3.1) is [1]

$$\sigma_{\text{Br}} = \frac{4}{3} \frac{A}{N_A X_0} \frac{1}{(\Delta p/p)_{\text{lim}}} \left[ \ln \frac{1}{(\Delta p/p)_{\text{lim}}} - \frac{5}{8} \right] \quad (27)$$

where  $A$  is the atomic weight of the scattering nucleus,  $N_A$  is Avogadro's number, and  $(\Delta p/p)_{\text{lim}}$  is the limiting momentum acceptance of the ring (either a longitudinal limit from the rf bucket height, or a transverse limit from scattering in a dispersive section of the ring). Note that, in a typical ring the residual gas pressure, and thus the decay rate, includes a term proportional to the beam intensity, leading to non-exponential decay. The decay rate for Touschek scattering also depends

on the beam intensity [2]. For typical collider parameters, beam-gas bremsstrahlung is the most important of the single-beam loss mechanisms. At highest energy circular  $e^+e^-$  colliders [4] the beam lifetime can be limited by beamstrahlung [5].

**Lifetime mitigation** For synchrotron light sources, it is desirable to keep a nearly constant current circulating in the ring to minimize cyclic thermal changes in precision optical components. For this reason, most modern light sources have adopted the technique of top-up injection [6], whereby small amounts of beam are injected on a quasi-continuous basis to keep the circulating current essentially constant. Especially for asymmetric colliders, an identical approach, which mitigates the decrease in luminosity due to current decay in one or both rings has been implemented [7, 8] (also see Sec.3.3.9.1). By making the average and peak beam currents more nearly the same, the gain in integrated luminosity can be substantial, up to 30–50%, and this has contributed to the world-record luminosity of  $2.1 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1}$  at KEKB [9].

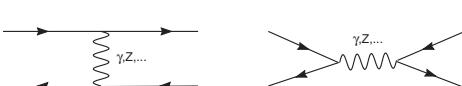
## References

- [1] F.C. Porter, NIM A302 (1991) 209
- [2] M.S. Zisman, J. Bisognano, S. Chattopadhyay, PA 23 (1988) 289; M.S. Zisman, J. Bisognano, S. Chattopadhyay, ZAP User's Manual, LBL-21270, UC-28 (1986)
- [3] J. Le Duff, NIM A239 (1985) 83
- [4] A. Blondel, F. Zimmermann, arXiv: 1112.2518 [hep-ex]
- [5] V. Telnov, arXiv: 1203.6563 [hep-ex]
- [6] H. Ohkuma, EPAC 08, 36, and references therein.
- [7] U. Wienands, PAC 05, 149
- [8] N. Iida et al, PAC 09, 2769
- [9] Y. Funakoshi et al, IPAC 10, 2372

### 3.3.3 Bhabha Scattering ( $e^+e^- \rightarrow e^+e^-$ )

*J.E. Spencer, SLAC*

The elastic scattering of electrons and positrons proceeds via one boson exchange and annihilation diagrams (the  $u$  and  $t$  channels for  $e^-e^- \rightarrow e^-e^-$ ):



Differential cross section with one photon exchange for unpolarized beams in CM [1]

$$\frac{d\sigma_{\text{Born}}^{\text{QED}}}{d\Omega}(e^+e^- \rightarrow e^+e^-) = \frac{\alpha^2 \hbar^2 c^2}{2s} \left[ \frac{1 + \cos^4 \frac{\theta}{2}}{\sin^4 \frac{\theta}{2}} - \frac{2 \cos^4 \frac{\theta}{2}}{\sin^2 \frac{\theta}{2}} + \frac{1}{2} (1 + \cos^2 \theta) \right]$$

when  $\sqrt{s} \equiv (p_1 + p_2)c \gg m_e c^2$  (typically  $s = 4E_b^2$ ).

The large forward cross section is useful for monitoring  $e^\pm$  luminosity while the larger angle scattering is used for electroweak measurements e.g. using the  $Z$  annihilation diagram. Based on lepton universality, the  $e^\pm$  annihilation diagram corresponds to  $e^+e^- \rightarrow \mu^+\mu^-$  and more generally, for elementary fermions  $f$  of charge  $Q_f e$ ,

$$\frac{d\sigma}{d\Omega}(l^+l^- \rightarrow f^+f^-) = \frac{\alpha^2 \hbar^2 c^2}{4s} (1 + \cos^2 \theta) Q_f^2$$

Integrating gives the (partial) total cross section of the expected luminosity or background rates in annular detectors. For example, for the u-quark with  $Q_f = 2/3$  and muons,

$$\sigma(e^+e^- \rightarrow u^+u^-) = \frac{4\pi\alpha^2 \hbar^2 c^2}{18s} \left[ \cos \theta + \frac{1}{3} \cos^3 \theta \right]_{\theta_2}^{\theta_1}$$

$$\sigma(e^+e^- \rightarrow \mu^+\mu^-) = \frac{9}{4} \sigma(e^+e^- \rightarrow u^+u^-)$$

where  $4\pi\alpha^2 \hbar^2 c^2 = 261 \text{ GeV}^2 \text{nb}$ .

**Møller scattering** The corresponding cross section for  $e^\pm e^\pm \rightarrow e^\pm e^\pm$  is [2]

$$\frac{d\sigma_{\text{Born}}^{\text{QED}}}{d\Omega}(e^\pm e^\pm \rightarrow e^\pm e^\pm) = \frac{\alpha^2 \hbar^2 c^2}{2s} \times \left( \frac{1 + \cos^4 \frac{\theta}{2}}{\sin^4 \frac{\theta}{2}} + \frac{1 + \sin^4 \frac{\theta}{2}}{\cos^4 \frac{\theta}{2}} + \frac{2}{\sin^2 \frac{\theta}{2} \cos^2 \frac{\theta}{2}} \right)$$

**Luminosity monitoring with Bhabha (Møller) scattering** BS is used to monitor or calibrate luminosity by selecting events with two oppositely charged tracks with high momentum (small  $p_t$  elastic) and small acollinearity angle that originate at the IP. Various contaminations such as muon pairs are statistically subtracted.

A quite good approximation for the singles rate (or pair coincidence rate) for an annular detector subtending an angular range  $\delta\theta$  ( $\theta_1$  to  $\theta_2$ ) near the IP – ignoring field effects from the detector or beam optics is

$$\sigma(\delta\theta) \approx \frac{4\pi\alpha^2 \hbar^2 c^2}{s} \left[ \frac{1}{\sin^2(\frac{\theta}{2})} + 4 \ln(\sin \frac{\theta}{2}) \right]_{\theta_2}^{\theta_1}$$

### Sec.3.3: PARTICLE-MATTER INTERACTION

For small angles,

$$\sigma(\delta\theta) \approx 4.2 \left( \frac{250 \text{ GeV}}{E_b} \right)^2 \frac{\theta_2^2 - \theta_1^2}{\theta_1^2 \theta_2^2} \text{ pb} \quad (6)$$

The expected Bhabha rate is given by  $\mathcal{L}\sigma(\delta\theta)|_{\theta_1}$ . If the predicted rate is reasonable, we can then consider the simulated backgrounds to estimate the signal-to-noise ratio. Increasing  $\theta_2$  is usually not effective.

It is fair to ignore the solenoid field if there is no crossing angle between the colliding beams, nor strong disruption from beam-beam effects. To account for beam-beam effects, solenoid fields, finite beam size and divergence effects there are Monte Carlo codes (e.g. ABEL, BHLUMI and others) that can be used to simulate the beams and their interactions at the IP for use with tracking codes such as MAD or TRANSPORT for optics design and electromagnetic shower codes such as GEANT or EGS4 for detector design and background simulation. (See also Sec.3.3.9.4.)

A semi-analytical way to use such codes is to transport the generated events (for the assumed conditions and perturbed by the beam-beam interaction, Sec.2.5) outside the primary detector. The resulting rate of scattered pair events and singles is then a function of the effective, disrupted luminosity, the acceptance of the detector and the intervening beam optics. A standard approximation, for the rate in a rectangular detector outside a solenoid, is

$$R_{\text{BS}}^{\text{QED}} = \mathcal{L} \frac{16\alpha^2 \hbar^2 c^2}{s} \int \int \frac{d\Omega}{\theta^4} \quad (7)$$

$$= \mathcal{L} \frac{16\alpha^2 \hbar^2 c^2}{s} \int_{x'} \int_{y'} \frac{dx' dy'}{(x'^2 + y'^2)^2}$$

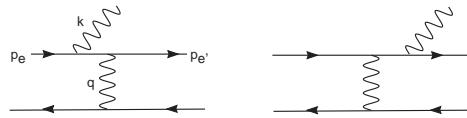
where  $\theta = \sqrt{x'^2 + y'^2}$ ,  $\tan\phi = y'/x'$ , and  $x'$ ,  $y'$  are the angles in cartesian coordinates.

For a transverse point source and dispersion free transport from the IP to the detector, the position of a particle at the detector is  $x_d = R_{12}x'$  and  $y_d = R_{34}y'$  where  $R_{12}$  and  $R_{34}$  can be calculated with an optics code, ( $R_{ij}$  = first order transport matrix elements, Sec.2.2.1). Substituting the BS detector's coordinates in the optical reference system gives the corresponding integration limits to compute the expected rates,

$$R_{\text{BS}}^{\text{QED}} = \mathcal{L} \frac{4\alpha^2 \hbar^2 c^2}{s} \left[ -\frac{1}{x' y'} \right]_{x'_1}^{x'_2} - \left( \frac{1}{x'^2} + \frac{1}{y'^2} \right) \tan^{-1} \left( \frac{y'}{x'} \right) \Big|_{y'_1}^{y'_2} \Big|_{x'_1}^{x'_2} \quad (8)$$

A detector with area  $L_x L_y$  might cover the region from  $x'_1 = X/R_{12}$  to  $x'_2 = (X + L_x)/R_{12}$  and from  $y'_1 = -L_y/(2R_{34})$  to  $y'_2 = L_y/(2R_{34})$ . The values of  $X$  (and/or  $Y$ ) and  $L_{x,y}$  are chosen to avoid the incoming and outgoing disrupted beams and backgrounds.

#### Radiative Bhabha scattering ( $e^+e^- \rightarrow e^+e^-\gamma$ )



A free, charged particle radiates when scattered or accelerated: "it never happens that a scattering event is unaccompanied by the emission of quanta" [3]. Pure Bhabha scattering occurs at extremely small and presumably unresolvable angles for leptons. For the RBS photons,  $\theta_k \sim \mathcal{O}(1/\gamma)$ . For the radiating lepton,

$$\theta_{l'} \sim \frac{1}{\gamma} \frac{1-w}{w} \sim \frac{1}{\gamma} \frac{E_k}{E_b - E_k} \quad (9)$$

where  $w = P_{e'}||/P_e \approx E_{e'}/E_b + \mathcal{O}(1/\gamma^2)$ . The RBS differential cross section in lowest order in the ultrarelativistic limit for the outgoing lepton is [4]

$$\frac{d\sigma_{e^-}}{dw} (e^\pm e^\mp \rightarrow e^\pm e^\mp \gamma) \quad (10)$$

$$= \frac{4\alpha r_e^2}{1-w} (1+w^2 - \frac{2}{3}w) \left( \ln \frac{s}{m_e^2} \frac{w}{1-w} - \frac{1}{2} \right)$$

and for the photons [4], with  $y = 1-w = E_k/E_b$ ,

$$\frac{d\sigma_\gamma}{dy} (e^\pm e^\mp \rightarrow e^\pm e^\mp \gamma) \quad (11)$$

$$= \frac{4\alpha r_e^2}{y} [y^2 + \frac{4}{3}(1-y)] \left( \ln \frac{s}{m_e^2} \frac{1-y}{y} - \frac{1}{2} \right)$$

where  $4\alpha r_e^2 = 2.32 \text{ mb}$ . Using a finite bunch size correction [5], a good approximation expected for the radiated photon power for some bandwidth  $\delta y$  into a forward detector is

$$P_\gamma \approx \mathcal{L} E_b \left( \frac{4}{3} \alpha r_e^2 \right) \left[ 2 + \ln \frac{\sigma_x \sigma_y}{\lambda_c (\sigma_x + \sigma_y)} \right] \times [y^3 - 2y^2 + 4y]_{y_{\min}}^{y_{\max}} \quad (12)$$

where  $\lambda_c = r_e/\alpha$  and  $\sigma$  is the rms beam size.

The value  $y_{\min} \rightarrow 0$  is never required in the above expressions because of cuts or detector resolution.

Some conclusions are:

- Elastic BS decreases inversely with  $s = 4E_b^2$ .
- RBS effects typically increase logarithmically with  $s/m^2$ .
- Photons are emitted within an angle  $\theta_\gamma = m/E_b$  with the particles.
- Their spectrum will run up to  $\sim E_b$  and vary as  $1/E_b$ .
- Hard scattering corresponds to angles outside the  $m/E_b$  cone.
- Typical beam divergence angles and luminosity monitors are  $\gg \theta_\gamma$ .
- Virtually all scattered particles that are detected are hard.
- Initial state radiation cannot be explicitly distinguished. This is true for the final state when the radiative event is untagged. This implies the outgoing scattered particles are not generally collinear. It also implies that the cm and lab systems are no longer the same.
- Beam-beam disruption will broaden the acollinearity distribution.
- Most experiments based on BS need to include radiative effects.

Electrodynamics suggests other radiative effects that need to be considered for BS in an external field or non-free environment. A more complete discussion is available [6] including comparisons to data, ranges of applicability of expressions and the importance of such radiative effects.

## References

- [1] H.J. Bhabha, Proc. Roy. Soc. 154 (1935) 195
- [2] C. Møller, Ann. d. Physik 14 (1932) 568
- [3] J. Schwinger, PR 75 (1949) 898
- [4] G. Altarelli, F. Buccella, Nuovo Cim. 34 (1964) 1337; G. Altarelli, B. Stella, Lett. Nuovo Cim. 9 (1974) 416
- [5] A.E. Blinov et al, PL B113 (1982) 423
- [6] J.E. Spencer, SLAC-ARDB Rept. (1997)

### 3.3.4 Compton Scattering ( $e^\pm\gamma \rightarrow e^\pm\gamma'$ )

*J.E. Spencer, SLAC*

Compton scattering (CS) [1] is a two-body, elastic process proceeding, in lowest order, via



The relativistic cross section for free electrons was given by Klein and Nishina [2]. Including helicities,  $\sigma_{e,e'} = \pm 1/2$ ,  $\lambda_{\gamma,\gamma'} = \pm 1$ , in the CM for  $\sqrt{s} \equiv (p_1 + p_2) c \gg m_e c^2$  gives

$$\frac{d\sigma_{\text{Born}}^{\text{QED}}}{d\Omega}(e^\pm\gamma \rightarrow e^\pm\gamma') \quad (1)$$

$$= \frac{\alpha^2 \hbar^2 c^2}{s} \begin{cases} \cos^{-2} \frac{\theta}{2} \sigma_{e,e'} = \pm \frac{1}{2}, \lambda_{\gamma,\gamma'} = \pm 1 \\ \cos^2 \frac{\theta}{2} \sigma_{e,e'} = \pm \frac{1}{2}, \lambda_{\gamma,\gamma'} = \mp 1 \end{cases}$$

CS of polarized laser photons on electron beams is used to measure e-beam polarization and provide higher energy polarized photons (Sec. 7.6.1).

### Classical Compton scattering ( $e^\pm + \omega \rightarrow e^\pm + \gamma$ )

CS as a frequency upshifting technique occurs in many ways, e.g. in wigglers, undulators and FELs where the initial photons are the low frequency, virtual components of static fields. Single-photon CS (order  $r_e^2$ ) in the laboratory is

$$\frac{d\sigma}{dy} = \frac{2\pi r_e^2}{x} \left[ (1-y) + \frac{1}{(1-y)} - \frac{4y}{x(1-y)} + \frac{4y^2}{x^2(1-y^2)} + 2\sigma_{e,e'} \lambda_{\gamma,\gamma'} F_P \right] \quad (2)$$

with  $F_P$  in terms of relativistic invariants  $x$  and  $y$

$$F_P = \frac{y}{(1-y)} (2-y) \left[ 1 - \frac{2y}{x(1-y)} \right]$$

and  $x$  and  $y$  in terms of 4-vectors  $p$  and  $k$

$$x = \frac{s}{p \cdot p} - 1 = \frac{2p \cdot k}{p \cdot p} = 2 \frac{E_b \omega}{m^2} (1 - \beta \cos \theta) \leq 0.0153 E_b [\text{GeV}] \omega [\text{eV}] \quad (3)$$

$\theta = 0$  defines head-on scattering in the lab,

$$y = 1 - \frac{u}{p \cdot p} = \frac{2p \cdot k'}{p \cdot p} = \frac{2E_b \omega'}{m^2} (1 - \beta \cos \theta') = \frac{\omega'}{E_b} + \mathcal{O}(\theta'^2) \quad (4)$$

$4\pi r_e^2 \approx 1$  barn. Integration gives the total cross section  $\sigma_C = \sigma_{NP} + 2\sigma_{e,e'} \lambda_{\gamma,\gamma'} \sigma_P$  where

$$\sigma_{NP} = \frac{2\pi r_e^2}{x} \left[ \left( 1 - \frac{4}{x} - \frac{8}{x^2} \right) \ln(1+x) + \frac{1}{2} + \frac{8}{x} - \frac{1}{2(1+x)^2} \right] \rightarrow \frac{2\pi r_e^2}{x} \left( \ln x + \frac{1}{2} \right) \quad (5)$$

$$\sigma_P = \frac{2\pi r_e^2}{x} \left[ \left( 1 + \frac{2}{x} \right) \ln(1+x) - \frac{5}{2} + \frac{1}{1+x} - \frac{1}{2(1+x)^2} \right] \rightarrow \frac{2\pi r_e^2}{x} \left( \ln x - \frac{5}{2} \right) \quad (6)$$

### Sec.3.3: PARTICLE-MATTER INTERACTION

The arrows give the ultrarelativistic case  $x \gg 1$ . For an electron at rest in the lab ( $x = 2\omega/m$ ),  $\sigma_{NP}$  and  $\sigma_P$  both fall rapidly compared to the Thomson limit for  $x > 1$ . For  $x \ll 1$ ,  $\sigma_{NP} = \frac{8\pi r_e^2}{3}(1-x)$ , the Thomson result. The total cross section typically depends less on the polarization than does the outgoing photon energy distribution.  $\sigma_P = 0$  for  $\ln x \approx 2.5$ , but for  $y = 0.7$  near  $y_{max} = x/(1+x)$ , the spectrum can differ by a factor of 6. The relation between incoming and outgoing photons is:

$$\frac{\omega_2}{\omega_1} = \frac{1 - \beta \cos \theta_1}{1 - \beta \cos \theta_2 + \frac{\omega_1}{\epsilon_1}(1 - \cos \phi)} \quad (7)$$

where  $\theta_1, \theta_2$  and  $\phi$  are the angles between the vectors  $(\vec{p}_1, \vec{k}_1)$ ,  $(\vec{p}_1, \vec{k}_2)$ , and  $(\vec{k}_1, \vec{k}_2)$  and  $\epsilon_1 = E_b$ . Fig.1 shows some characteristic distributions.

**Multiphoton Compton scattering** ( $e^\pm + n\omega \rightarrow e^\pm + \gamma$ ) Intense laser fields modify the expressions above. Multiphoton effects can appear in many ways e.g. with a laser having  $n_\gamma \lambda^3 > 1$  where  $n_\gamma = I/(c\hbar\omega)$  is the number density of nearly monochromatic photons. Using  $I = E^2/377\Omega$  for 1 eV photons shows that present terawatt tabletop lasers greatly exceed this. One effect is to increase the outgoing photon energy. This regime can be characterized by two dimensionless, classical and quantal strong-field invariants:  $\Upsilon$  (Beamstrahlung parameter, Sec.2.5.3.2) involving a particle's Compton wavelength and  $\eta$  involving the photon's wavelength,

$$\Upsilon = e\lambda_C \left[ \frac{|\langle (F_{\mu\nu} p_1^\nu)^2 \rangle|}{(p_1 \cdot p_1)^2} \right]^{\frac{1}{2}} \equiv \frac{E_{rms}^*}{E_c} \quad (8)$$

$$\eta = \frac{e\sqrt{|\langle A_\mu A^\mu \rangle|}}{m_e c^2} = \frac{eE_{rms}}{m_e c^2} \lambda = \Upsilon \frac{m_e^2 c^4}{2\hbar\omega_1 \epsilon_1}$$

$A_\mu$  is the 4-vector potential of the incident field.  $E_{rms}^*$  is the rms electric field in the electron rest frame.  $\eta=1$  ( $\Upsilon=1$ ) corresponds to an energy gain of one electron mass over one photon (Compton) reduced wavelength  $\lambda$  ( $\lambda_C$ ).  $E_c$  is called the “critical field” of QED,

$$E_c = \frac{m^2 c^3}{e\hbar} = 1.3 \times 10^{16} \text{ V/cm} \quad (9)$$

While it is not possible to create static, critical fields in the lab frame, highly relativistic electrons channeling through a crystal or electrons and positrons at the IP of the next-generation linear colliders may experience such fields from their opposing beams (Sec.2.5.3.2).

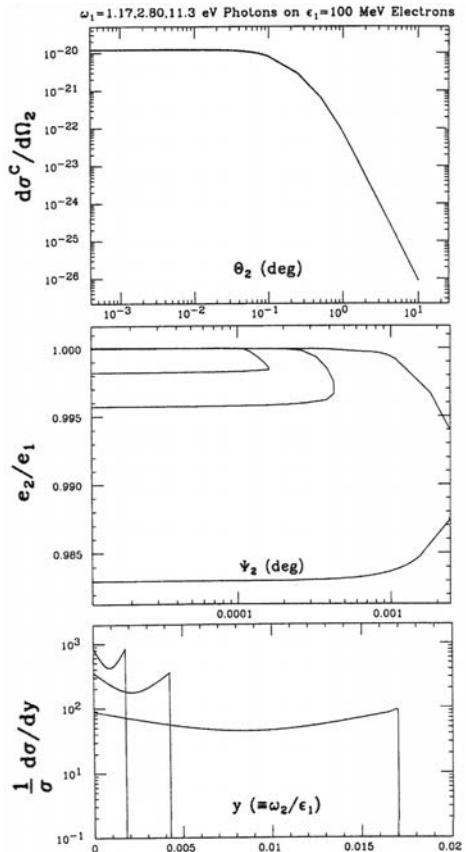


Figure 1: Unpolarized Compton distributions for the outgoing photons and electrons in the lab.

An electron in the wave field behaves as if it has an effective mass

$$\bar{m}_e^2 c^4 = \epsilon^2 - (p_\parallel c)^2 = m_e^2 c^4 (1 + \eta^2) \quad (10)$$

This effect is identifiable by a shift in the kinematic edge for CS and the term  $\eta^2 m_e c^2 / 2$  can be identified as the “ponderomotive potential” in the equivalent Hamiltonian for the problem. The effective mass of the electron in the strong field is taken into account by replacing  $p^\mu$  by the “quasimomentum”  $q^\mu$  (and  $p'^\mu \rightarrow q'^\mu$ ),

$$q^\mu = p^\mu + \frac{\eta^2 m^2}{2(k \cdot p)} k^\mu \quad (11)$$

The kinematics of the scattering process are given by  $q^\mu + nk^\mu = q'^\mu + k'^\mu$ . Here  $n$  is the number of absorbed photons and

$$\epsilon_{\text{edge}}(n, \eta) = \frac{\epsilon_1}{1 + 2n(k \cdot q)/m^2} \quad (12)$$

The fact that the kinematic edge decreases with increasing  $n$  is essential in distinguishing electrons scattered via  $n > 1$  processes from ordinary  $n = 1$  CS. Nonlinear CS has been observed [3].

The transition probability for nonlinear Compton scattering is [4]

$$\begin{aligned} \frac{dW}{du} &= \sum_{n=1}^{\infty} \frac{dW_n}{du} \\ &= \sum_{n=1}^{\infty} \frac{2\pi r_0^2}{u_1} \left\{ -\frac{4}{\eta^2} J_n^2(z) + \left( 2 + \frac{u^2}{1+u} \right) \right. \\ &\quad \times \left. [J_{n-1}^2(z) + J_{n+1}^2(z) - 2J_n^2(z)] \right\} \end{aligned}$$

where

$$u = \frac{k \cdot k'}{k \cdot q'}, u_1 = \frac{2k \cdot q}{m^2}, z = \frac{2\eta\sqrt{u(u_n - u)}}{u_1\sqrt{1+\eta^2}}$$

and  $u_n = nu_1$ . An expansion, for small values of  $\eta$ , shows that the contributions from the  $n^{\text{th}}$  order multiphoton process scale as  $\eta^{2n}$ .

**Compton scattering on background radiation** Feenberg and Primakoff (1948) first considered the scattering of fast electrons on various sources of cosmic radiation. With the discovery of 2.7 K thermal background radiation, others considered how this restricted the range and energy of different particles that are observable in the vicinity of the earth. Later, Telnov and Dehning et al. [5] considered how scattering from thermal background photons can limit storage ring lifetimes and increase detector backgrounds. This is a straightforward application of above formulae and the Planck distribution for spectral density of thermal photons,

$$\frac{dn_\gamma}{d\omega} = \left( \frac{\omega^2}{\pi^2(\hbar c)^3} \right) \frac{1}{e^{(\frac{\omega}{kT})} - 1} \quad (13)$$

assumed to be isotropic in space. Number density  $n_\gamma = 20.2 T^3 \text{ cm}^{-3}$  with average photon energy  $\bar{\omega} = 2.7 k_B T$ . The electron energy loss distribution for a beam of  $N$  particles over a distance  $L$  is

$$\frac{dN}{dy} = NL \int \int \frac{d\sigma}{dy} (1 + \cos \theta_1) dn_\gamma(\omega, T) \frac{d\Omega}{4\pi} \quad (14)$$

This can also be integrated over  $y \leq y_{\text{max}} = x/(1+x)$ . Using  $\bar{\omega}$  in the expression for  $x$  and averaging over incoming angles gives (see also Eq.(15), Sec.4.6)

$$\frac{\Delta N}{N} \approx \frac{1}{2} n_\gamma L \sigma_C(\bar{x}) \quad (15)$$

## References

- [1] A.H. Compton, PR 21 (1923) 715
- [2] O. Klein, Y. Nishina, Z. Physik 52 (1929) 853
- [3] C. Bula, et al, PRL 76 (1996) 3116
- [4] N.B. Narozhnyi, A.I. Nikishov, V.I. Ritus, JETP 20 (1965) 622
- [5] V.L. Telnov, NIM A260 (1987) 304; B. Dehning et al, PL B249 (1990) 145

### 3.3.5 Limit of Focusing of Electron Beam due to Synchrotron Radiation

K. Oide, KEK

**Symbols** Assume focusing in one plane ( $y$ ). Incoming rms invariant emittance  $\gamma \varepsilon_y$ , focusing strength and length of the final lens  $K = B'/(B\rho)$ ,  $L$ . Courant-Snyder parameters at the IP  $\beta_y^*, \alpha^* = 0$ . Distance between the FP and the final lens  $\ell^* \gg \beta_y^*$ . Unperturbed orbit at the IP  $(y_0^*, y'_0)$ . Deviation of the orbit at the IP due to radiation  $Y$ . Photon energy emitted in the lens,  $u(s)$ . Expected number of photons per unit length of the lens  $N(s)$ . Fourier transform of photon spectrum  $S(P)$ .

**Minimum rms spot size** The rms spot size at the FP is [1]

$$\begin{aligned} \sigma_y^{*2} &= \beta_y^* \varepsilon_y + \frac{110}{3\sqrt{6\pi}} r_e \lambda_e \\ &\quad \times F(\sqrt{KL}, \sqrt{K}\ell^*) \left( \frac{\gamma^2 \varepsilon_y}{\beta_y^*} \right)^{5/2} \end{aligned} \quad (1)$$

The second term represents the blow-up due to energy fluctuation caused by synchrotron radiation in the final lens. Dimensionless  $F(\sqrt{KL}, \sqrt{K}\ell^*)$  depends on the configuration of the focusing (Fig.1),

$$\begin{aligned} &F(\sqrt{KL}, \sqrt{K}\ell^*) \\ &\equiv \int_0^{\sqrt{KL}} d\phi \left| \sin \phi + \sqrt{K}\ell^* \cos \phi \right|^3 \\ &\quad \times \left[ \int_0^\phi \left( \sin \phi' + \sqrt{K}\ell^* \cos \phi' \right)^2 d\phi' \right]^2 \end{aligned} \quad (2)$$

### Sec.3.3: PARTICLE-MATTER INTERACTION

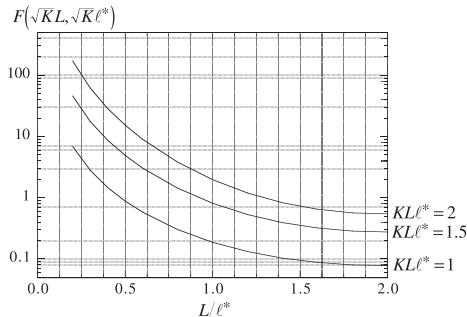


Figure 1: Function  $F(\sqrt{KL}, \sqrt{K\ell^*})$ .

The minimum rms spot size

$$\sigma_y^* \min = \sqrt{\frac{7}{5}} (\gamma \varepsilon_y)^{5/7} \quad (3)$$

$$\times \left[ \frac{275}{3\sqrt{6\pi}} r_e \lambda_e F(\sqrt{KL}, \sqrt{K\ell^*}) \right]^{1/7}$$

is achieved when

$$\beta_y^* = \left[ \frac{275}{3\sqrt{6\pi}} r_e \lambda_e F(\sqrt{KL}, \sqrt{K\ell^*}) \right]^{\frac{2}{7}} \gamma (\gamma \varepsilon_y)^{\frac{3}{7}} \quad (4)$$

In [1], the horizontal beam size  $\sigma_{x'}^*$  was ignored. A modified expression including a correction can be found in [2]. The additional contribution from  $\sigma_{x'}^*$  can be made small by decreasing the strength of the horizontally focusing magnet in the final focus lens.

**Reduction of luminosity** Since the resulting distribution of the beam differs from Gaussian (unperturbed core, plus long tails), the luminosity is better than that calculated from Eq.(3). The minimum effective beam size determined by luminosity is typically  $\frac{1}{2}\sigma_y^* \min$ .

Luminosity  $\mathcal{L}$  is reduced from the value  $\mathcal{L}_0$  without radiation in the lens by the factor [3]

$$D = 2\sqrt{\frac{\beta_y^* \varepsilon_y}{\pi}} \int_0^\infty dk \exp(-\beta_y^* \varepsilon_y k^2) \times \langle |\tilde{\rho}_Y(k)| \rangle_{y_0^*, y_0'^*}^2 \quad (5)$$

where  $\langle |\tilde{\rho}_Y(k)| \rangle_{y_0^*, y_0'^*}$  is the Fourier transform of the distribution of  $Y$ , averaged over  $(y_0^*, y_0'^*)$ , and

$$|\tilde{\rho}_Y(k)| = \exp \left[ \int_0^L N(s)(S(P) - 1) ds \right]$$

$$N(s) = \frac{5\alpha\gamma}{2\sqrt{3}} |Kg(s)y'_0|$$

$$S(P) = \frac{3(\sqrt{1+P^2} + P)^{5/3}}{10P\sqrt{1+P^2}}$$

$$- \frac{3(\sqrt{1+P^2} - P)^{5/3}}{10P\sqrt{1+P^2}}$$

$$P = \frac{3}{2} \lambda \gamma^2 k y'_0 K^2 G(s) |g(s)|$$

$$g(s) = \frac{\sin \sqrt{K}s}{\sqrt{K}} + \ell^* \cos \sqrt{K}s$$

$$G(s) = \int_0^s g(s)^2 ds$$

$$Y \approx Ky_0'^* \int_0^s G(s) \frac{u(s)}{E} ds$$

### References

- [1] K. Oide, PRL 61 (1988) 1713
- [2] J. Irwin, Eq.(11.111), SLAC-474 (1996)
- [3] K. Hirata, B. Zotter, K. Oide, PL B224 (1989) 437

### 3.3.6 Thermal Outgassing and Beam Induced Desorption

V. Baglin, E. Mahner, CERN  
O. Gröbner, ret. CERN

#### Thermal outgassing rates [1, 2]

$Q$  (Torr l s<sup>-1</sup> cm<sup>-2</sup>)

$$Q = C \exp(-E_b/k_B T)$$

depends strongly on surface coverage  $C$ , temperature  $T$  and molecular binding energy  $E_b$  [3]. Molecules may exist with different binding states between 0.8 and 2 eV [4]. A low outgassing rate requires a chemical cleaning procedure adapted to the material [5]. Coverage is also determined by diffusion from the bulk, a particularly strong effect for hydrogen in most metals [6]. For clean, unbaked Al, Cu and stainless steel (SS) surfaces after 100 h pumping at room temperature, the main gas species desorbed are fairly similar, Tab.1.

The quoted rates may decrease by several orders of magnitude with pumping time as lower binding energy states empty. In unbaked systems water dominates and  $Q$  remains sensitive to temperature [7]. H<sub>2</sub>O is found to decreases approximately inversely with time

$$Q_{H_2O} \sim Q_o t^{-1}$$

Table 1: Unbaked outgassing rates  $Q$ .

| Gas                  | $Q$ (Torr $1\text{ s}^{-1}$ $\text{cm}^{-2}$ ) |
|----------------------|--|
| $\text{H}_2$         | $7 \cdot 10^{-12}$                             |
| $\text{CH}_4$        | $5 \cdot 10^{-13}$                             |
| $\text{H}_2\text{O}$ | $3 \cdot 10^{-10}$                             |
| CO                   | $5 \cdot 10^{-12}$                             |
| $\text{CO}_2$        | $5 \cdot 10^{-13}$                             |

Table 2: Outgassing rates (Torr  $1\text{ s}^{-1}$   $\text{cm}^{-2}$ ) for Cu, Al, and SS after a bakeout followed by 50 hours of pumping.

| Gas           | Cu                 | Al and SS          |
|---------------|--------------------|--------------------|
| $\text{H}_2$  | $1 \cdot 10^{-12}$ | $5 \cdot 10^{-13}$ |
| $\text{CH}_4$ | $5 \cdot 10^{-15}$ | $5 \cdot 10^{-15}$ |
| CO            | $1 \cdot 10^{-14}$ | $1 \cdot 10^{-14}$ |
| $\text{CO}_2$ | $5 \cdot 10^{-15}$ | $1 \cdot 10^{-14}$ |

For chemically cleaned and baked surfaces ( $150^\circ\text{C}$  for Al and Cu,  $300^\circ\text{C}$  for SS), and following 50 h pumping, outgassing rates are dominated by  $\text{H}_2$ , Tab.2. Within practical limits, it is preferable to increase temperature instead of extending baking time.  $\text{H}_2$  outgassing of SS below  $10^{-13}$  Torr  $1\text{ s}^{-1}$   $\text{cm}^{-2}$  can be obtained by firing the steel at  $950^\circ\text{C}$  in a vacuum furnace.

At cryogenic temperatures below 20 K, thermal outgassing vanishes for all species apart from  $\text{H}_2$  and He.

**Photon induced gas desorption at room temperature** The desorption yield  $\eta = \text{desorbed molecules/incident particle}$ . The yield for photons varies with gas species, photon energy, photon angle of incidence, surface cleanliness and temperature. With increasing photon dose,  $D$ , the surface becomes cleaner and the initial  $\eta_0$  decreases at a rate which may be different for different species, see Fig.1 for baked Cu [8].  $D$  is expressed in photons/m of irradiated vacuum chamber since in an accelerator the photons impinge on part of the vacuum chamber wall and the remaining surface is irradiated by scattered and reflected photons. The main species desorbed are  $\text{H}_2$ ,  $\text{CH}_4$ , CO and  $\text{CO}_2$ . Even with baked chambers,  $\text{H}_2\text{O}$  and  $\text{O}_2$  are released as the surface carbon concentration is reduced. For baked SS  $\eta_0$  is similar to Cu; for Al it is about a factor 10 higher. For activated, TiZrV coated chambers  $\eta_0$  is 2 orders of magnitude less compared to Cu. Molecular desorption by photons is generally attributed to a two-

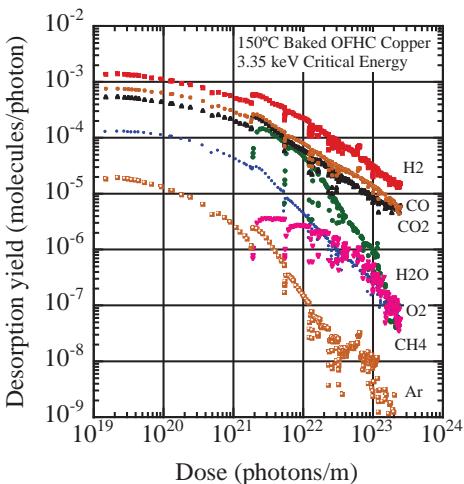


Figure 1: Desorption yield for baked OFHC Cu at 3.35 keV critical energy versus dose.

step process, whereby first photoelectrons are produced, which excite molecules bound in near surface layers. Subsequently, excited more weakly bound molecules desorb thermally. The thermal desorption may persist long after the direct photon irradiation has stopped. Conversely, it takes time to re-establish a quasi-steady rate after a stop.

As a function of dose  $D$  (photons/m)

$$\eta = \eta_0 D^{-\alpha} \text{ where } \alpha \sim 0.6 \text{ to } 1.$$

As a function of  $Q$  (Torr l/m)

$$\eta = \eta_0 \exp(-Q/Q_0).$$

Values for baked OFHC Cu are listed in Tab.3.

In a circular machine photons originate predominantly in bending magnets. As function of critical energy,  $E_c$ ,  $\eta$  scales roughly like  $\eta \sim E_c^\alpha$  with  $\alpha \sim 0.7$  to 1.2 below 1 keV. Fig.2 has been compiled from different measurements for Cu and Al over a wide range of  $E_c$ . At large  $D$  ( $> 10^{23}$  photons/m), the surface becomes very clean and a pumping effect of  $\sim 100$  l/s/m for CO,  $\text{O}_2$  and  $\text{CO}_2$  has been measured while the surface re-saturates with gas. Cu desorption yields of less than  $10^{-6}$  molecules/photon have been measured at 11 keV critical energy with an accumulated dose of  $2 \times 10^{25}$  photons/m [9].

**Photon induced gas desorption at low temperature** At 77 K and 4.2 K the same species  $\text{H}_2$ ,  $\text{CH}_4$ , CO and  $\text{CO}_2$  are desorbed as at RT. Figure 3 shows  $\eta$  of Cu at 77 K, 10 mrad grazing incidence

Table 3: Values of  $\eta_0$  and  $Q_0$  for baked OFHC Cu.

|                             | $H_2$               | $CH_4$              | $CO$                | $CO_2$              |
|-----------------------------|---------------------|---------------------|---------------------|---------------------|
| $\eta_0$ (molecules/photon) | $9.2 \cdot 10^{-4}$ | $2.3 \cdot 10^{-4}$ | $3.7 \cdot 10^{-4}$ | $5.5 \cdot 10^{-4}$ |
| $Q_0$ (Torr l/m)            | $3.0 \cdot 10^{-2}$ | $4.5 \cdot 10^{-4}$ | $8.4 \cdot 10^{-3}$ | $1.1 \cdot 10^{-2}$ |

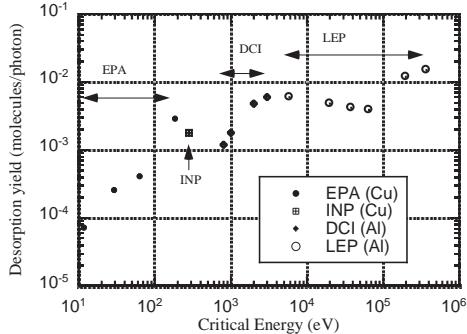
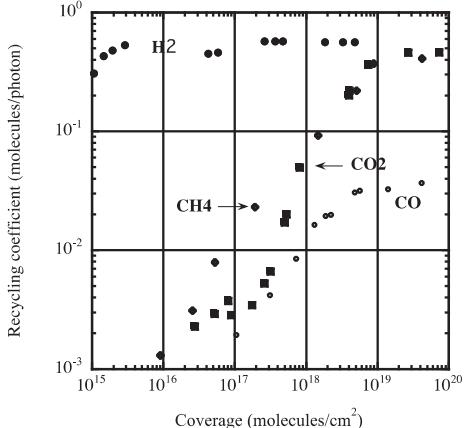
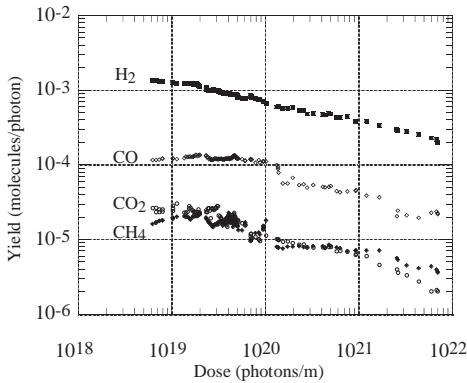
Figure 2: Desorption yield versus  $E_c$  compiled from different experiments.

Figure 4: Recycling of weakly bound, physisorbed molecules versus coverage.

Figure 3: Desorption yield of Cu at 77 K, 10 mrad grazing incidence and  $E_c = 50$  eV.

and  $E_c = 50$  eV [10]. At 4.2 K  $\eta_0$  is reduced by about one order of magnitude compared to room temperature and  $\alpha \sim 1/3$  [11].

At low temperature the surface is ‘sticky’ since the mean sojourn time  $\tau(s) \approx 10^{-13} \exp(E_b/k_b T)$  becomes very long [12]. Thus, under photon irradiation, molecules desorb initially from a strongly bound state and accumulate in a weakly bound physi-sorbed state ( $E_b \sim 20$  to 100 meV). As the surface coverage,  $\Theta$  (molecules/cm<sup>2</sup>) increases,

the photon induced recycling,  $\kappa$ , of physisorbed molecules becomes important, particularly for  $H_2$ , see Figure 4 [13]. Condensed molecules such as  $CH_4$  and  $CO_2$  may also be cracked into  $H_2$ ,  $CO$  and  $O_2$  by photons [14].

In a cold vacuum system external pumps are inefficient since longitudinal gas flow is strongly suppressed. In the LHC, molecules are effectively pumped on the 1.9 K magnet cold bore. Molecules can reach the cold bore through perforations in the beam screen, i.e., a fraction,  $f$ , of the specific surface area,  $F$ . The volume density  $n$  (molecules/cm<sup>3</sup>) and surface coverage  $\Theta$  on the screen held at 5 to 20 K interact through mutual desorption and re-adsorption [15]

$$V \frac{dn}{dt} = q - an + b\Theta \quad \text{and} \quad F \frac{d\Theta}{dt} = cn - b\Theta .$$

Here  $q = \eta\Gamma$  is the desorption rate of strongly bound molecules by the photon flux,  $\Gamma$ .

$b = F/\tau + \kappa\Gamma$  describes physisorbed molecules, which desorb spontaneously after a sojourn time  $\tau$ , as well as by photons.

$c = v_m sF/4$  gives the wall pumping with the sticking probability  $s$ , and the mean velocity for molecules with mass  $m$ ,  $v_m = \sqrt{8k_b T/\pi m}$ .

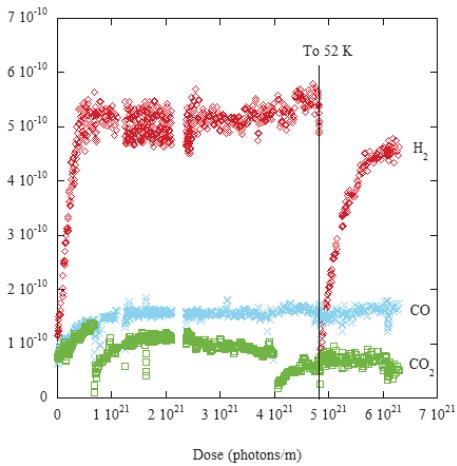


Figure 5: Pressure response to synchrotron radiation in a perforated Cu-coated SS liner at 10–20 K inside a cold bore at  $\sim$ 3 K.

$a = c + \nu_m F f / 4$  represents the surface pumping and the pumping by the perforation in the beam screen. The equilibrium gas density  $n_{equ} = \frac{4\eta\Gamma}{v_m F f}$  depends only on the desorption yield, which decreases gradually with photon dose, and on the pumping through the beam screen perforations.  $\Theta$  and  $\kappa$  determine the time response to reach the steady state. The pressure response to synchrotron radiation in a test system for the LHC is shown in Figure 5. Temporary warming to 52 K can be used to remove the physisorbed  $H_2$  layer.

**Desorption by residual gas ions** With a positively charged beam of current  $I$ , residual gas ions are accelerated to the wall and cause ion induced desorption [16]. The flux of desorbed molecules

$$Q(I) = \eta\sigma_i PI/e + Q_0.$$

For relativistic particles the ionisation cross-section  $\sigma_i \sim 10^{-18} \text{ cm}^2$ . With the thermal outgassing  $Q_0$  and the linear pumping speed  $S (\text{l s}^{-1} \text{ m}^{-1})$  the dynamic pressure is

$$P(I) = Q(I)/S = Q_0/(S - \eta\sigma_i I/e).$$

The vacuum stability limit is given by  $(\eta I)_{\text{crit}} = Se/\sigma_i$ . During operation, when the product  $\eta I$  approaches the stability limit, the local pressure increases drastically and ultimately leads to beam loss [17].

$\eta$  is a function of gas species, ion energy and surface cleanliness. Measured yields for unbaked SS are shown in Fig.6.

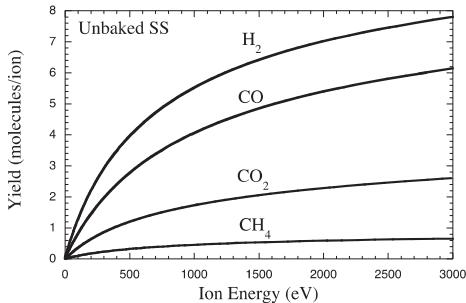


Figure 6: Ion induced desorption yield for unbaked SS as a function of ion impact energy.

The yields for baked SS are about 5 times lower. Since ions are produced from the residual gas,  $\eta$  is a net yield, as ions have a probability to be implanted in the surface. In a clean vacuum chamber, the overall effect may become negative resulting in ‘beam-pumping’.

To improve surface cleanliness many cleaning methods (bakeout, Ar-O<sub>2</sub> glow discharge, surface coating) can be used.  $S$  is ultimately limited by the specific molecular conductance  $c_m (\text{m}^4 \text{s}^{-1})$  of the vacuum pipe. In a room-temperature system with regularly spaced, ‘lumped’ pumps (of speed  $2S$ ) at distance  $2L$  the limit is

$$(\eta I)_{\text{crit}} < \pi^2 / 4(e/\sigma_i)(c_m/L^2).$$

For a uniform circular pipe  $c_m = (2\pi/3)v_m r^3$ .

The conductance limitation can be overcome with linear, distributed pumping. With the large molecular mass, CO becomes the most critical gas. Since the residual gas is a mixture of  $H_2$ ,  $CH_4$ , CO and  $CO_2$  the single gas stability limit can be significantly reduced by the desorption of one species by another. Thus an  $H_2$  dominated gas composition will change to CO and the simple expression may overestimate the stability limit.

In a cold-bore vacuum system the stability limit is given by the wall-pumping  $(\eta I)_{\text{crit}} = (\nu_m/4)(e/\sigma_i)sF$ .

For the LHC and a molecular sticking probability  $s \sim 1$ ,  $(\eta I)_{\text{crit}} \sim 5 \times 10^3 \text{ A}$ . This effect is partly offset by  $\eta$  which increases approx. proportionally to the surface coverage;  $\eta > 10^3$  molecules/ion at a monolayer of condensed  $H_2$  [18]. To guarantee stability, a minimum pumping speed is provided by the perforations in the beam screen, which ensures a coverage below a monolayer and  $(\eta I)_{\text{crit}} \sim 10^2 \text{ A}$ .

Table 4:  $\eta$  for differently prepared targets bombarded with  $1.5 \cdot 10^9 \text{ Pb}^{53+}$  ions at 14 mrad grazing incidence.

| Material   | Vac. firing | Treatment                                    | Coating                         | Bakeout                  | $\eta$ [mol/ion] |
|------------|-------------|--|---------------------------------|--------------------------|------------------|
| SS (316LN) | 950°C, 2h   | -  | -                               | 300°C, 24h               | $2.0 \cdot 10^4$ |
| SS (316LN) | 950°C, 2h   | Ar-O <sub>2</sub> glow disch.                | -                               | 350°C, 24h               | $6.2 \cdot 10^4$ |
| SS (316LN) | 950°C, 2h   | -  | TiZrV (1.5 $\mu\text{m}$ sput.) | 200°C, 24h               | $2.7 \cdot 10^2$ |
| SS (316LN) | 950°C, 2h   | -  | TiZrV (1.5 $\mu\text{m}$ sput.) | 300°C, 24h               | 59               |
| SS (316LN) | 950°C, 2h   | Ar-O <sub>2</sub> glow disch.                | -                               | 350°C, 24h               | $2.9 \cdot 10^4$ |
| SS (316LN) | 950°C, 2h   | 50 $\mu\text{m}$ electropol.                 | -                               | 300°C, 24h               | $1.0 \cdot 10^4$ |
| SS (316LN) | 950°C, 2h   | 150 $\mu\text{m}$ electropol.                | -                               | 400°C, 24h               | $9.3 \cdot 10^3$ |
| SS (316LN) | 950°C, 2h   | -  | Au (30 $\mu\text{m}$ galvanic)  | 300°C, 24h               | $1.4 \cdot 10^3$ |
| SS (316LN) | 950°C, 2h   | 50 $\mu\text{m}$ chem. pol.                  | -                               | 300°C, 24h               | $6.5 \cdot 10^3$ |
| SS (316LN) | 950°C, 2h   | Liner: Si coated SS                          | Si (0.4 $\mu\text{m}$ evap.)    | 300°C, 24h               | $3.5 \cdot 10^4$ |
| SS (316LN) | 950°C, 2h   | He-O <sub>2</sub> glow disch.                | -                               | 350°C, 24h               | $1.7 \cdot 10^4$ |
| SS (316LN) | 950°C, 2h   | Liner: 127 $\mu\text{m}$ Mo foil             | -                               | 300°C, 24h               | $2.4 \cdot 10^4$ |
| SS (316LN) | 1050°C, 2h  | 50 $\mu\text{m}$ chem. pol.                  | -                               | 300°C, 24h               | $6.2 \cdot 10^3$ |
| SS (316LN) | 1050°C, 2h  | Vented after scrubb.                         | -                               | 300°C, 24h               | $5.3 \cdot 10^3$ |
| SS (316LN) | 1050°C, 2h  | -  | Ag (2 $\mu\text{m}$ galvanic)   | 300°C, 24h               | $1.2 \cdot 10^3$ |
| SS (316LN) | 1050°C, 2h  | 50 $\mu\text{m}$ chem. pol., getter purified | -                               | 300°C, 24h               | $6.9 \cdot 10^3$ |
| SS (316LN) | 950°C, 2h   | Liner: St707 getter strips                   | -                               | 350°C, 24h<br>+400°C, 3h | $2.5 \cdot 10^2$ |
| SS (316LN) | 950°C, 2h   | 150 $\mu\text{m}$ electropol.                | Pd (0.6 $\mu\text{m}$ sput.)    | 300°C, 24h               | $2.6 \cdot 10^2$ |
| SS (304L)  | -           | -  | -                               | 300°C, 24h               | $3.2 \cdot 10^3$ |
| Cu (OFE)   | -           | -  | -                               | 150°C, 24h               | $2.2 \cdot 10^4$ |
| Al (6082)  | -           | -  | -                               | 150°C, 24h               | $2.1 \cdot 10^4$ |

**Heavy-ion induced gas desorption** Large pressure rises by several orders of magnitude can be caused by lost beam ions that impact under grazing angle onto vacuum chamber walls. The heavy-ion induced gas desorption can seriously limit the ion intensity, luminosity and beam lifetime, as observed at CERN, GSI, and BNL [19].

Measurements of  $\eta$  (molecules/ion) using 4.2 MeV/u lead ions from CERN LINAC3, bombarding differently prepared accelerator-type vacuum chambers under 14 mrad grazing incidence angle, are summarized in Tab.4 [20, 21]. The desorption yield depends critically on the surface properties of the baked chambers. For bare SS  $\eta$  varies between 3000 and  $2 \times 10^4$  molecules/ion. Noble metal or getter coatings (Au, Ag, Pd, TiZrV) reduce  $\eta$  by a factor of up to 100. Glow discharged (Ar-O<sub>2</sub>, He-O<sub>2</sub>) SS, Al, Cu, and Mo have  $\eta \geq 10^4$  molecules/ion. The *ex situ* vacuum firing temperature (950°C or 1050°C) of a SS substrate, the thickness of a noble metal coating, and

the *in situ* bakeout temperature (300°C or 400°C) have no significant influence on  $\eta$ . Nitrogen venting of a heavy-ion scrubbed SS surface to atmosphere results in the complete loss of the cleaning effect. Perpendicular lead ion impact results in the lowest  $\eta$ , which increases for grazing angles in the mrad range. The dominant gases desorbed by lead ions are CO, CO<sub>2</sub>, and H<sub>2</sub>. A clear correlation exists between the surface oxygen and carbon content and the measured  $\eta$ .

Heavy ion-induced  $\eta$ , measured as a function of the ion energy and comprising different types of ions, charge states, target types, materials, and impact angles are summarized in Fig.7, where molecular desorption yields derived from dedicated beam line experiments (HCX: K<sup>+</sup>, LINAC 3: Pb<sup>53+</sup>, HLI: Pb<sup>27+</sup>, Zn<sup>10+</sup>, Xe<sup>18...21+</sup>, TSL: Ar<sup>8...12+</sup>, HHT: Ar<sup>10+</sup>, U<sup>73+</sup>, SPS: In<sup>49+</sup>) are compared with results obtained from machine experiments (AGS: Au<sup>31+</sup>, SIS 18: U<sup>28+</sup>, RHIC: U<sup>73+</sup>, Cu<sup>29+</sup>). The yield scales with the

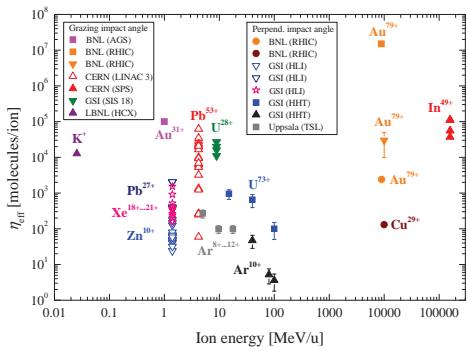


Figure 7: Heavy-ion induced desorption yields classified by experiments with either grazing or perpendicular ion-impact angles [19].

electronic energy loss of the projectile [22]

$$\eta = k \left( \frac{dE}{dx_{el}} \right)^n .$$

For Ar and U ions (5–100 MeV/u), impacting on SS targets, one finds  $k_{\text{Ar}} = 1.7$ ,  $k_{\text{U}} = 0.9$ ,  $n_{\text{Ar}} \approx 2.1$ , and  $n_{\text{U}} \approx 2.9$  [23];  $n > 2$  indicates a thermally moderated desorption process [24].

Mitigation techniques to overcome high  $\eta$  values were first applied in LEIR at CERN, which is equipped with NEG-coated SS vacuum chambers and gold-coated absorbers to catch charge-exchanged lead ions under perpendicular impact.

## References

- [1] CAS, CERN 99-05 19 (1999)
- [2] CAS, CERN-2007-003 (2007)
- [3] A.G. Mathewson et al, 7th Int. Vacuum Congress (1977) 1027
- [4] M. Moraw, Vacuum 36, 7-9 (1986) 523
- [5] CAS, CERN 94-01 v 2 (1994) 535
- [6] M.D. Malev, Vacuum 23, 2 (1973) 43
- [7] A. Berman, Vacuum 47, 4 (1996) 327
- [8] O. Gröbner et al, JVST A 12(3) (1994) 846
- [9] K. Kanazawa et al, NIM A499 (2003) 66
- [10] R. Calder et al, JVST A 14(4) (1996) 2618
- [11] V. Baglin et al, Vacuum 67 (2002) 421
- [12] CAS, CERN-2004-008 (2004) 396
- [13] V.V. Anashin et al, Vacuum 53, 1-2 (1999) 269
- [14] V.V. Anashin et al, Vacuum 60, 1-2 (2001) 15
- [15] O. Gröbner, Vacuum 60, 1-2 (2001) 25
- [16] I.R. Collins et al, e<sup>+</sup>e<sup>-</sup> Factories'99, KEK Proceedings 99-24 (2000) 88
- [17] O. Gröbner, R. Calder, PAC 73, (1973) 760
- [18] J.C. Barnard et al, Vacuum 47, 4 (1996) 347

- [19] E. Mahner, PRST-AB 11, 104801 (2008)
- [20] E. Mahner et al, PRST-AB 6, 013201 (2003)
- [21] E. Mahner et al, PRST-AB 8, 053201 (2005)
- [22] A. Molvik et al, PRL 98, 064801 (2007)
- [23] H. Kollmus et al, JVST A27 (2009) 245
- [24] K. Wien, Radiation Effects and Defects in Solids 109 (1989) 137

## 3.3.7 Photoemission and Secondary Emission

V. Baglin, CERN

### 3.3.7.1 Photoemission

Photoemission in a circular machine is mainly due to bending magnet radiation. Photoemission is a 3-step phenomenon [1]. First, a photon is absorbed inside the material over a few optical attenuation lengths (~5 nm for metallic surfaces at 10 mrad incidence angle). Second, an electron is excited and diffuses through the solid while losing energy and creating secondary electrons by inelastic collisions during the process. Third, the electrons escape from the solid across the potential barrier. The escape depth of the excited electrons is roughly independent of the materials. Electrons with kinetic energies up to 1 keV have escape depth of the order of 3-5 nm. Photoemission is therefore surface sensitive. When a photon irradiates a surface, a photoelectron can be emitted if the photon energy,  $h\nu$ , is larger than the work function,  $W_F$ , of the material. The kinetic energy of the emitted electrons varies from 0 to  $(h\nu - W_F)$ . Most of the photoelectrons are secondary electrons with low energies and only 0.1-1% have high energies. Fig.1 shows the energy distribution curves (EDC) of evaporated gold when irradiated by photon energies of 35 and 100 eV. Electrons originating from the valence band and core levels can be identified from the EDCs. X-ray photoelectron spectroscopy (XPS) is widely used in surface science to identify the components on the material's surface.

When irradiated by synchrotron radiation, the vacuum chamber is subjected to irradiation ranging from a few tenth of eV to a few keV. Therefore, the produced photoelectron spectrum is the sum of the monochromatic EDCs in the range of interest. This spectrum is dominated by secondary electrons whose energy is below 10 eV.

The photoelectron yield, PY\*, which is the ratio of photoelectrons per absorbed photon, is a function of the material, its surface cleanliness, its surface treatment, and of the incidence angle.

### Sec.3.3: PARTICLE-MATTER INTERACTION

Table 1: Forward scattering photon reflection and photoelectron yield for materials subjected to synchrotron radiation with 45 and 194 eV [3].

| Material            | Status             | 45 eV |            | 194 eV |            |
|---------------------|--------------------|-------|------------|--------|------------|
|                     |                    | R (%) | PY* (e/ph) | R (%)  | PY* (e/ph) |
| Al                  | unbaked            | -     | 0.11       | -      | 0.32       |
| Cu-smooth           | unbaked            | 81    | 0.11       | 77     | 0.32       |
| Cu-electrodeposited | unbaked            | 5     | 0.08       | 7      | 0.08       |
| Cu-sawtooth         | unbaked            | 8     | 0.03       | 7      | 0.04       |
| TiZr                | unbaked            | 20    | 0.06       | 17     | 0.08       |
| TiZr                | activated at 350°C | 20    | 0.02       | 17     | 0.03       |

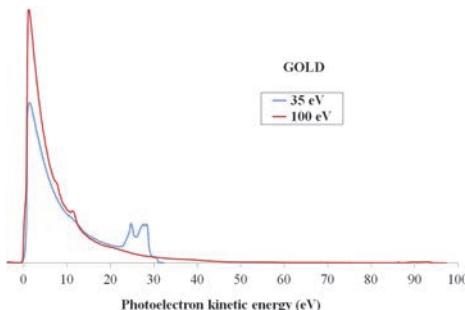


Figure 1: Photoelectron spectra of gold when irradiated by photon energies of 35 and 100 eV [2].

Since the photoelectron yield measurements of interest are performed at grazing angle on samples of a few meters in order to mimic the machine operation condition, the forward reflectivity of the photons, R, must be taken into account to derive PY\*. Tab.x1 shows typical results measured in a photon beam line when a vacuum chamber is subjected to synchrotron radiation at grazing angle (11 mrad). For smooth Cu, at 4 keV critical angle, R and PY\* equals 33% and 0.43 e/ph.

When a “sawtooth structure” is impressed on the horizontally outward side of the beam pipe [3], the photoelectron yield and the forward reflectivity are reduced to a few %. Photon conditioning also decreases PY\* [3].

Since the photoelectrons spectrum is dominated by low energy electrons, a magnetic field of about 50 G parallel to the emitting surface can also be used to significantly reduce the photoelectron emission [4].

In a cryogenic environment, the photoelectron yield is slightly reduced. The effect of condensed gas is negligible up to 10 monolayers.

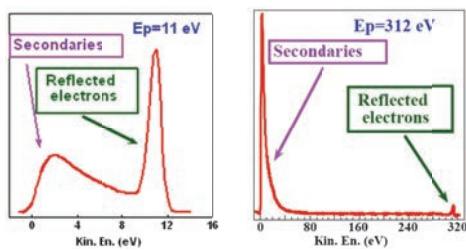


Figure 2: EDCs of Cu at 10 K when bombarded by electrons with energies of 11 and 312 eV [6].

#### 3.3.7.2 Secondary emission

Primary electrons impinging on a surface cause the emission of secondary electrons. Secondary emission is a 3 steps process. First, a primary electron of energy  $E_P$  penetrates into the solid and generates secondary electrons. Second and third, secondary electrons diffuse through the solid and are evacuated into the vacuum [5]. The penetration depth of the primary electrons is described by the stopping power law. It increases with the electron kinetic energy and is in the range of 1-5 nm for 200-500 eV electrons. The emitted electrons are divided into 3 populations. The true secondary electrons have a kinetic energy of less than 50 eV. The elastically reflected primary electrons have a kinetic energy  $E_P$ . The inelastically backscattered electrons have kinetic energies from 50 eV to  $E_P$ . Auger electrons with defined peak energy can also be visible in the EDC. Fig.2 shows an example of two EDCs of Cu at 10 K when bombarded by electrons with primary energies of 11 and 312 eV. The ratio of reflected electrons to the true secondary electrons decrease with the increase of the kinetic energy of the primary electrons. Above 50 eV, this ratio is less

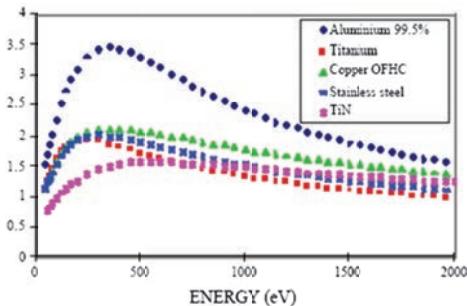


Figure 3: SEY of some technical surfaces [7].

than 10% whereas the contribution of the reflected electrons is dominant at very low energy.

The secondary electron yield (SEY),  $\delta$ , is defined as the ratio of the number of produced electrons to the number of incident electrons. Fig.3 shows the SEY curves for different technical surfaces measured at perpendicular incidence. The maximum of the SEY, ( $\delta_{\max}$ ), is assumed at a certain value of the primary incident electron energy, called  $E_{\max}$ , which typically is in the range 200-300 eV. The SEY of technical surfaces can differ greatly from an atomically clean surface due to the presence of a surface layer of oxide and contaminants: sputter cleaned Cu has a  $\delta_{\max}$  of 1.3 whereas as received Cu has  $> 2$ . The SEY of Cu is the same at room temperature and at cryogenic temperature. However, the condensation of gas affects the SEY: 200 monolayers of water on sputter cleaned Cu increase  $\delta_{\max}$  to 2.3.

The SEY curve can be described by the sum of true secondary electrons and reflected electrons [6, 8]:

$$\delta(E_P) = \delta_{\text{true}}(E_P) + R\delta_{\text{elastic}}(E_P)$$

where

$$\delta_{\text{true}}(E_P) = \delta_{\max} \frac{s \times (E_P/E_{\max})}{s - 1 + (E_P/E_{\max})^s}$$

with  $s \sim 1.4$ . When decreasing the incidence angle, the SEY increases. Introducing the angle  $\theta$  with respect to the surface normal, the parameters  $\delta_{\max}$ ,  $E_{\max}$  and  $s$  may vary like [8, 9]:

$$\delta_{\max}(\theta) = \delta_{\max} e^{0.4(1-\cos\theta)}$$

$$E_{\max}(\theta) = E_{\max} \times (1 + 0.7(1 - \cos\theta))$$

$$s(\theta) = s \times (1 - 0.18(1 - \cos\theta))$$

The azimuthal distribution of the true secondary electrons follows the Beer-Lambert law. The reflection of the elastic electrons is specular.

One mitigation of the electron cloud in accelerator is obtained by thin films coatings. These coatings are applied onto the vacuum chamber wall to reduce  $\delta$ . TiN, activated TiZrV and unbaked amorphous carbon coatings have respectively a  $\delta_{\max}$  of 1.5, 1.1 and 1.0. Beside a surface treatment which reduces the oxide layer, the surface morphology also plays an important role and allows further reduction of the SEY below unity. Production of macroscopic grooves is another mean to reduce the SEY [7, 10].

In accelerator machines, external means to reduce and control the SEY such as magnetic fields, clearing electrodes and beam conditioning are also used. A 50 G magnetic field parallel to the electron emitting surface reduces the emission of the true secondary electrons, allowing the mitigation of the electron cloud by solenoid magnets in field free regions. In superconducting cavity rf couplers, a bias voltage of 2.5 kV is often applied on clearing electrodes to avoid multipacting. Beam conditioning with electrons is used to reduce the SEY of technical surfaces such as copper and stainless steel. During this process, a graphitization of the oxide carbon is observed and  $\delta_{\max}$  is typically reduced to 1.1-1.3 after a dose of 10 mC/mm<sup>2</sup> [11]. The conditioning efficiency of low energy electrons with less than 50 eV is weaker than for higher energy electrons. Some beam conditioning is also observed under photon irradiation.

## References

- [1] C.N. Berglund et al, PR 136, A1030 (1964)
- [2] R. Cimino et al, PRST-AB 2, 063201 (1999)
- [3] I.R. Collins et al, EPAC 98; Y. Suetsugu et al, JVST A 21 (1) (2003) 186
- [4] V.V. Anashin et al, Vacuum 60 (2001) 225
- [5] H. Bruining. Physics and applications of secondary electron emission, Pergamon Press (1954); O. Hachenberg et al, Adv. Electron Phys. 11 (1959) 413-499
- [6] R. Cimino et al, App.Surf.Sci. 235 (2004) 231
- [7] N. Hilleret et al, EPAC 00, 217
- [8] M.A. Furman, CERN LHC Project Report 180 (1998) or LBNL-41482/CPB note 247; G. Ru-molo et al, PRST-AB 4, 012801 (2001)
- [9] R.E. Kirby, F.K. King, NIM A469 (2001) 1
- [10] I. Montero et al, App.Surf.Sci. 253 (2007) 7627; M. Pivi et al, NIM A551 (2005) 187
- [11] C. Scheuerlein et al, App.Surf.Sci. 202 (2002) 57; M. Nishiwaki et al, ECLOUD 07

### Sec.3.3: PARTICLE-MATTER INTERACTION

#### 3.3.8 Ionization Processes

*F. Zimmermann, CERN*

Cross section for *collisional ionization* is [1]

$$\sigma_{\text{coll}} \approx 4\pi \left(\frac{\lambda_e}{\beta}\right)^2 \left[ C_1 + 2C_2 \left( \ln \beta\gamma - \frac{1}{2} \right) \right] \quad (1)$$

where  $C_1$  and  $C_2$  depend on gas properties, and  $v = \beta c$  is the velocity of the ionizing charged particle. Assuming the latter is a singly charged  $e^-$  or proton, for carbon monoxide  $C_1 = 35$  and  $C_2 = 3.7$ , and  $\sigma_{\text{coll}} \approx 2$  Mbarn, at highly relativistic energies ( $\gamma = 10^3$ – $10^6$ ). The ion density per meter  $\lambda_{\text{ion}}$  after the passage of  $N$  beam particles is (at 300°K)

$$\lambda_{\text{ion}} [\text{m}^{-1}] = \sigma_{\text{ion}} \rho_m N \approx 6 N P_{\text{gas}} [\text{Torr}] \quad (2)$$

where  $P_{\text{gas}}$  gas pressure, and  $\rho_m$  molecular density. In future accelerators, *tunneling ionization* may be significant; the corresponding ionization rate is approximately [2, 3]

$$W \approx 8 \frac{\alpha^2 c}{\lambda_e^2} \frac{E_{\text{ion}}}{e\mathcal{E}} \exp \left( -\frac{4}{3} \frac{\alpha}{\lambda_e} \frac{E_{\text{ion}}}{e\mathcal{E}} \right) \quad (3)$$

where  $\mathcal{E} \approx e\lambda_{\text{beam}}/[2\pi\epsilon_0(\sigma_x + \sigma_y)]$  the collective bunch field and  $E_{\text{ion}}$  the ionization energy. Tunneling ionization thresholds measured in experiments on beam-driven plasma wake-field acceleration [4] are consistent with first-order calculations of the “ADK” approximation for the field-ionization rate [5].

In  $e^\pm$  storage rings, many ions are produced via *photoionization* [6]. Photoionization cross section for hydrogen is [7, 6]

$$\sigma_{\text{pi}} \approx 5.35 \times 10^{-20} \left( \frac{100}{\epsilon} \right)^{3.228} (\text{cm}^2) \quad (4)$$

$$\text{for } \epsilon > \epsilon_{\text{th}} \approx 18 \text{ eV}$$

and for carbon monoxide

$$\sigma_{\text{pi}} \approx 9.12 \times 10^{-14} \epsilon^{-2.48} - 4.80 \times 10^{-12} \epsilon^{-4.05} (\text{cm}^2) \quad \text{for } \epsilon > \epsilon_{\text{th}} \approx 14 \text{ eV} \quad (5)$$

where  $\epsilon$  is the photon energy and  $\epsilon_{\text{th}}$  the threshold energy, both in eV. Photoionization cross sections for many elements are tabulated in [8]. Inside the beam, collisional ionization produces about 3 times more ions than photoionization. Most photoionization processes occur between beam radius and vacuum chamber. The total number of ions generated by photoionization is typically 2–4 times that produced by collisional ionization. Average ion production rate (ions produced per second and per electron) due to photoionization is

$$\frac{1}{\tau_{\text{pi}}} = l_i \rho_m \int_0^C ds \int_{\epsilon_{\text{th}}}^{\infty} \sigma_i(\epsilon) n(\epsilon, s) d\epsilon \quad (6)$$

where  $C$  is the ring circumference or beam-line length,  $l_i \approx \sqrt{2b\rho}$  is the average path of photons until they hit the chamber ( $\rho$  is the bending radius,  $b$  the horizontal chamber half size), and  $n(\epsilon)d\epsilon$  is the number of quanta emitted per unit time per electron with energies between  $\epsilon$  and  $\epsilon + d\epsilon$ , evaluated at location  $s$ . If only ions inside the rms beam size are of interest, replace  $b$  by  $\sigma_x$  in the argument of  $l_i$ .

For the lifetime of ion beams, ion-atom impact ionization cross sections [9] are important.

**Limits on ion accumulation** [10, 11] Accumulation of ions is limited by: (i) residual-gas density  $d_{\text{gas}}$  and multiple ionization ( $d_{\text{ion}} \leq d_{\text{gas}}$ ); (ii) neutralization [ $d_{\text{ion}} \leq n_B N_B / (2\pi\sigma_x\sigma_y C)$ ]; (iii) heating by beam-gas collisions; (iv) ion drifts or recombination with photoelectrons. Accumulation limit in undulators is not clear [6, 11].

#### References

- [1] F.F. Rieke, W. Prepejchal, PR A6 (1972) 1507
- [2] T.O. Raubenheimer, PAC 95, 2752
- [3] L.D. Landau, E.M. Lifshitz, Quantum Mechanics: Non-Relativistic Theory, Pergamon (1981)
- [4] C.L. O'Connell et al, PRST-AB 9, 101301 (2006)
- [5] M.V. Ammosov et al, JETP 64 (1986) 1191
- [6] Y. Miyahara, Jpn. J. Appl. Phys. 26, no.9 (1987) 1544
- [7] E.L. Kasarev, E.R. Dodolyak, Opt. Spectrosk. 56 (1984) 643
- [8] D.E. Cullen et al, UCRL-50400-Vol.6-Rev.4-Pt.A, DE90 010470 (1989)
- [9] I.D. Kaganovich et al, NJP 8 (2006) 278
- [10] A. Poncet, Lecture Notes in Phys. 400, Springer Verlag (1994)
- [11] Y. Baconnier, A. Poncet, P. Tavares, CERN/PS 94-40 (1994)

#### 3.3.9 Beam Induced Detector Backgrounds and Irradiation in $e^+e^-$ Colliders

*S.D. Henderson, FNAL  
M. Sullivan, SLAC*

Consideration of detector backgrounds and irradiation is important in modern  $e^+e^-$  colliders which require unprecedented colliding beam currents and precision charged particle tracking detectors near the IP [1, 2, 3, 4, 5, 6, 7, 8]. High current  $e^-p$  colliders also fall into this category. A masking system is incorporated to shield the

detector from beam-generated background. Collimators are required upstream of the IP to remove errant beam particles or to shield the detector from synchrotron radiation produced by particles at high transverse amplitudes. The degree of background suppression is set by detector tracking and trigger requirements, and by component lifetime. Interaction region shielding designs and operational performance for the B-factories are discussed in [9] for PEP-II, [10] for KEK-B and [11] for CESR. The BEPCII  $e^+e^-$  collider is described in [12], and the Super B-factory designs in [13] for the Italian SuperB and in [14] for Super-KEKB.

### 3.3.9.1 Sources of detector backgrounds

Detector backgrounds arise from the following processes: (i) the interaction of beam particles with residual gas (beam-gas backgrounds), (ii) synchrotron radiation (SR) generated in nearby magnets, (iii) Touschek scattering, (iv) beam-beam interactions, (v) interaction with thermal photons, and (vi) operational particle losses due to injection, tuning, beam loss, etc. In general, (i)-(v) may be estimated through simulation, while (vi) is difficult to estimate.

**Beam-gas interactions** The most important beam-gas interactions are Coulomb scattering and bremsstrahlung. These interactions produce beam particles with large transverse oscillation amplitudes that may follow trajectories which strike the vacuum chamber near the IR, generating shower debris which reaches the detector.

Lost particle and energy fluxes We estimate the particle flux in a “target” region (the interaction region mask) extending from  $s_1$  to  $s_2$ , due to beam-gas interactions at  $s_0$  ( $s_0 < s_1 < s_2$ ). Let the Coulomb scattering angle required to just strike the aperture at  $s_1$  be  $\theta_1$  and that required to just strike at  $s_2$  be  $\theta_2$  ( $\theta_1 > \theta_2$ ). For bremsstrahlung, the electron energy loss,  $\delta = (E - E_0)/E_0$ , required to strike the aperture at  $s_1$  is  $\delta_1$ , and that required to strike at  $s_2$  is  $\delta_2$ , where  $-1 < \delta_1 < \delta_2 < 0$ . These limits may be easily determined with a particle tracking code that includes full off-momentum beam transport capability. Typical limits for the energy loss range from many times the beam energy spread to several percent. The flux of particles striking the “target” region from a source region between  $s_0$  and  $s_0 + ds$  is

$$dN(s_0) = I_{\text{beam}} \sigma_T(s_0) \mathcal{N}(s_0) ds \quad (1)$$

where  $\mathcal{N}$  is the gas density and  $\sigma_T = \sigma_T^{\text{Coul}} + \sigma_T^{\text{Brem}}$  is the total cross-section folded with the transport efficiency from source to target. If the transport efficiency from source to target is 100% [compare Eq.(6), Sec.3.3.3],

$$\sigma_T^{\text{Coul}}[\text{barns}] = \frac{0.2606 Z^2}{E^2[\text{GeV}^2]} \left( \frac{1}{\theta_2^2} - \frac{1}{\theta_1^2} \right) \frac{1}{[\text{mrad}^2]} \quad (2)$$

$$\begin{aligned} \sigma_T^{\text{Brem}}[\text{barns}] &= 0.00309 Z(Z+1) \ln \frac{183}{Z^{1/3}} \\ &\times \left[ \ln \left( \frac{\delta_1}{\delta_2} \right) + (\delta_1 - \delta_2) + \frac{3}{8}(\delta_1^2 - \delta_2^2) \right] \end{aligned} \quad (3)$$

Finally,

$$\frac{dN[\text{particle/s}]}{ds[\text{cm}]} = 202 \sigma_T[\text{barn}] I[\text{A}] P[\text{nTorr}] \quad (4)$$

The power striking the target region is

$$\begin{aligned} dP[\text{GeV/s}] &= 202 I[\text{A}] P[\text{nTorr}] ds[\text{cm}] E_0[\text{GeV}] \\ &\times \left[ \sigma_T^{\text{Coul}}[\text{barn}] + \int_{\delta_1}^{\delta_2} (\delta + 1) \frac{d\sigma^{\text{Brem}}}{d\delta} \epsilon(\delta) d\delta \right] \end{aligned} \quad (5)$$

where  $\epsilon(\delta)$  is the transport efficiency. It should be emphasized that the residual gas pressure has both a static and current dependent component.

Coulomb scattering (Sec.3.3.1) Consider an elliptical vacuum chamber having walls at  $\pm H(s)$  in  $x$  and  $\pm V(s)$  in  $y$ . A Coulomb scattering at  $s_0$  with scattering angle  $\theta$  and azimuthal angle  $\phi$  gives a trajectory  $x(s) = x_c(s) + S_x \theta \cos \phi$  and  $y(s) = y_c(s) + S_y \theta \sin \phi$  where  $x_c(s)$ ,  $y_c(s)$  is the closed orbit, and  $S_x = \sqrt{\beta_x(s)\beta_x(s_0)} \sin \psi_x(s - s_0)$ . The scattering angle required to strike the vacuum chamber at  $s$  (where a collimator or mask is located) is given by solution of

$$\begin{aligned} V^2(s)[x_c(s) + S_x \theta \cos \phi]^2 \\ + H^2(s)[y_c(s) + S_y \theta \sin \phi]^2 = H^2(s)V^2(s) \end{aligned} \quad (6)$$

(The trajectory may strike the vacuum chamber upstream of  $s$ .) Scatters  $\pi/2, 3\pi/2, \dots$  in betatron phase from the IP may be regions with large Coulomb contribution.

Bremsstrahlung (Sec.3.3.1) results in an off-energy particle (with energy error  $\delta$ ) and a photon (with energy  $|\delta|E$ ). The scattering angle of order  $1/\gamma$  may generally be neglected.

Scatters located near  $\pi, 2\pi, \dots$  in horizontal betatron phase from the IP may be regions with large bremsstrahlung contribution. Additionally, large amplitudes result when the source or target are in a region of high dispersion.

### Sec.3.3: PARTICLE-MATTER INTERACTION

A bremsstrahlung interaction at  $s_0$  gives  $x(s) = x_c(s) + \delta[D(s) - C_x D(s_0) - S_x D'(s_0)]$ , where  $C_x = \sqrt{\beta_x(s)/\beta_x(s_0)}[\cos \psi_x(s - s_0) + \alpha_x(s_0) \sin \psi_x(s - s_0)]$ . The vertical amplitude remains unchanged (assuming no  $y$  dispersion). The particle strikes the chamber wall at  $s$  when  $x^2(s) = H^2(s)$ . The energy loss required to strike the chamber at  $s$  is

$$\delta_{\pm}(s) = \frac{\pm H(s) - x_c(s)}{D(s) - C_x D(s_0) - S_x D'(s_0)} \quad (7)$$

where  $+$  is for an “outside” hit and  $-$  for an “inside.” (There is only a single physical solution since  $-1 < \delta < 0$ .)

Eq.(7) is valid in linear approximation only. For  $\delta \lesssim -0.01$ , one has to include higher order chromatic effects. Such a treatment is easily performed with first order matrix multiplication, taking into account the energy loss by proper adjustment of magnet strengths.

**Synchrotron radiation backgrounds** SR (Sec.3.1) generated in bends or quads (due to finite beam size or displaced trajectories) near the IP strikes the vacuum chamber and detector shielding and may be scattered into the detector beam pipe by any of the usual x-ray scattering mechanisms (Rayleigh scattering, Compton scattering, or x-ray fluorescence; Secs.3.3.1, 3.3.4). In addition, x-rays generated from beam particles at large amplitudes may directly strike the detector beam pipe.

Fig.1 shows the various transport mechanisms that give rise to SR backgrounds and dose. The mechanisms are grouped into three categories: (i) direct SR, (ii) backscattering either from the nearest masking surfaces or from downstream vacuum chamber surfaces, and (iii) forward scattering either from mask tips (tip scattering) or from the upstream vacuum chamber surfaces.

To calculate SR fluxes, we need to know: (i) the number of photons emitted in path length  $ds$ ,

$$\frac{dn}{ds} \left[ \frac{\text{photons}}{\text{m-A-s}} \right] = \frac{5}{2\sqrt{3}} \frac{\alpha\gamma}{\rho e} = 1.29 \times 10^{20} \frac{E[\text{GeV}]}{\rho[\text{m}]} \quad (8)$$

and (ii) the normalized x-ray spectrum (Sec.3.1.3),

$$\frac{1}{n} \frac{dn}{dE} = \frac{8}{15\sqrt{3}} \frac{1}{E} S \left( \frac{E}{E_c} \right) \quad (9)$$

For estimates of SR scattering in the range of interest here,  $E_\gamma \lesssim 100$  keV, realistic scattering

cross sections must be used. The Rayleigh scattering cross-section is given by

$$d\sigma_R/d\Omega = (d\sigma_T/d\Omega)[F(x, Z)]^2$$

where  $F(x, Z)$  is the atomic form-factor [15] as a function of momentum transfer  $x$ , and atomic number  $Z$ , and  $\sigma_T$  is the Thomson cross-section. The Compton scattering cross-section is similarly modified,

$$d\sigma_C/d\Omega = (d\sigma_{KN}/d\Omega)[S(x, Z)]$$

where  $S(x, Z)$  is the incoherent scattering function [16], and  $\sigma_{KN}$  is the Klein-Nishina cross-section (Sec.3.3.1). Total cross-sections are tabulated in [17]. The photoabsorption cross-section goes as  $Z^5/E^{7/2}$ , forward Rayleigh as  $Z^2$ , and the Compton as  $Z$ . Therefore, the ratio of total absorbed to scattered radiation goes as  $\sim Z^3$ .

The critical energy  $k_c$  of the photon energy spectrum of a bending magnet is

$$k_c[\text{keV}] = 0.666 E^2 [\text{GeV}] B [\text{T}]$$

Although half of the total energy of the photon spectrum is above this number only 8.7% of the total photons are above this energy where the photon flux is falling exponentially. The photon spectrum from focusing elements is more complicated but the spectrum generally falls exponentially due to most of the critical energy values being small and because the beam particle density is falling off rapidly as the transverse position of the beam particle increases.

Some of the SR backgrounds arise from scattering mechanisms in the extreme forward ( $\theta \lesssim 5^\circ$ ), or backward ( $\theta \gtrsim 175^\circ$ ) directions.  $F(x, Z)$  has a maximum at  $x = 0$  (forward scattering), where  $F(0, Z) = Z$ , and decreases to zero for large  $x$  (backscattering of high-energy x-rays).  $S(0, Z) = 0$ , and increases to  $S \simeq Z$  for large  $x$ . As a result, forward scattering is dominated by Rayleigh scattering. Backscattering of low-energy x-rays is dominated by Rayleigh scattering, and backscattering of high-energy x-rays is dominated by Compton scattering. The two backscattering cross-sections are approximately equal for  $E_\gamma \simeq 20$  keV in Cu, and  $E_\gamma \simeq 40$  keV in Ta. The function  $S$  is always less than  $Z$ , so that the real bound electron Compton scattering is always less than the KN formula. In the backward direction,  $S = 0.9Z$  at 30 keV in Cu, and 60 keV in Ta.

Photoabsorption followed by x-ray fluorescence contributes to both forward and back scattering, since photoemission is isotropic. K-shell

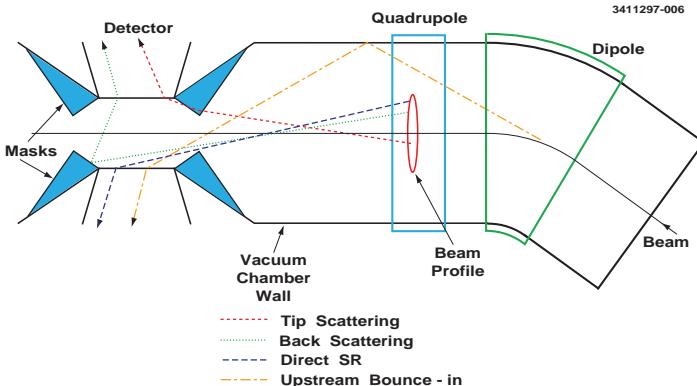


Figure 1: Synchrotron radiation scattering processes.

Table 1: K-shell photoelectric data.

| Element | K edge<br>keV | $\mu_{\text{photo}}/\rho$<br>$\text{cm}^2/\text{g}$ | $\omega_K$ | $E_K$<br>keV |
|---------|---------------|---|------------|--------------|
| Al      | 1.56          | 4890  | 0.039      | 1.49         |
| Cu      | 8.98          | 300   | 0.440      | 8.14         |
| Ag      | 25.51         | 56.4  | 0.831      | 22.59        |
| W       | 69.53         | 10.8  | 0.958      | 60.94        |
| Au      | 80.73         | 8.46  | 0.964      | 70.69        |

edge energies, total photoelectric mass attenuation coefficients, fluorescence yields  $\omega_K$  [18] and weighted K x-ray energies are given for a few elements in Tab.1 [17]. The L-shell emission must be taken into account for high- $Z$  materials. The L-shell fluorescence yields are much smaller; for Ta they are  $\sim 0.25$ . Absorption in the L edge can be an effective technique for reducing characteristic x-ray emission. If the K-shell energy is above the critical energy of the photon spectrum the material is considered a good absorber of the incident spectrum.

For tip scattering, only interactions within a few  $\lambda_{\text{mfp}}$  of the tip contribute to the scattering; tip scattering rates are dominated by x-rays which strike within  $\sim 10 \mu\text{m}$  of the mask tip.

**Beam-beam interactions** Detector “backgrounds” may arise during  $e^+e^-$  collisions from ordinary QED processes. These scale with luminosity rather than beam current. For example, radiative Bhabha scattering (Sec.3.3.3) was an important process at PEP-II due to the presence of the dipole located 20 cm from the IP [19, 20]. BELLE, the KEKB detector, essentially did not

see this background due to the outgoing beams having a first bend in an offset trajectory in the shared final focus magnet that is nearly 2 m from the IP. A comparison of B-factory IR designs can be found in [21]. Both very high luminosity Super-B factory designs must keep beam bending in the IR to a minimum in order to minimize this background source. Another QED process that becomes important at high luminosity is  $e^+e^-$  pair production. These low energy particles are generally wrapped up by the detector magnetic field, but the process can set a lower limit on the beam pipe radius at the collision point.

**Beamstrahlung** At proposed highest-energy circular  $e^+e^-$  colliders (e.g. LEP3, TLEP, SuperTRISTAN) beamstrahlung (Sec.2.5.3), generating off-energy particles, may become an important background source [22].

**Touschek scattering** Touschek scattering is a source of off-energy beam particles arising from the elastic scattering of particles within a bunch (Sec.2.4.12). Scattering results in two particles (with energy errors  $+\delta$  and  $-\delta$ ) which follow cosine-like betatron trajectories about the off-energy closed orbit. General techniques for shielding off-momentum particles generated by bremsstrahlung are applicable to the Touschek background. The Touschek lifetime scales as  $\gamma^2$ . So it is a concern in low-energy colliders such as DAΦNE [23] and CESR-C [24], where a set of movable collimators is (was) installed well away from the detector. New collider designs and new light source rings use very low emittance beams to achieve desired performance, so that Touschek

### Sec.3.3: PARTICLE-MATTER INTERACTION

lifetimes and backgrounds have become important design criteria.

**Thermal photon compton scattering** An inelastic scattering process resulting in off-energy beam particles (Sec.3.3.4), this is important at very high energy colliders (LEP, LEP2) [25, 26].

**Injection losses** Radiation during injection can arise from either the injected charge or from the stored beam. The injected charge may generate radiation that could be more effectively transported to the detector by virtue of the large oscillation amplitude. The manipulation of the stored beam may generate SR or lost particle backgrounds. Experience at CESR with two different IR configurations has shown that radiation during injection accounted for  $\sim 50\%$  of the total IR radiation dose for CESR Phase I operation and  $\sim 20\%$  for Phase II [5, 7].

**Continous injection** Both B-factories developed methods of continuously injecting charge into the storage rings while the detector is operational [27, 28]. The injection rates were usually 3-10 Hz depending on stored beam lifetime. Injecting while the detector was taking data was achieved by careful study of the detector backgrounds on a pulse by pulse basis looking for correlations between large background pulses and errant injection bunches. Light source facilities now have adopted continuous injection (or top-up) as a requisite of the accelerator design. A nice summary of light source top-up was given by Ohkuma [29] (also see Sec.3.3.2.2).

**Ion trapping** Electron beam destabilization due to trapped ions (Sec.2.4.13) leads to poor stored  $e^-$  lifetimes and may produce sudden large detector background rates, depending on the source location. B-factory designs incorporate a clearing gap in the bunch train to reduce this effect [30]. Electron-cloud effects (Sec.2.4.14) on the positron beam usually affect the luminosity of the collider before the detector sees an increase in background levels. However, in the p-p collider LHC, electron-cloud build up leads to higher detector background through increased vacuum pressure [31].

#### 3.3.9.2 Detector and IR radiation tolerance and budget

A detector background budget is based on three factors: (i) detector component radiation sensitivity, (ii) detector occupancy, and (iii) trigger rates.

The following are radiation dose limits for common detector/IR components:

**Crystal Scintillators** CsI, CsI(Tl): 1-100 krad, depending on dopant, impurities, manufacturer [32, 33, 34]. For other crystals, see [35].

**Drift Chamber:** 0.5 C/cm accumulated on sense wire [36, 37, 38]

**Silicon Detectors:**  $\sim 1$  Mrad ( $e^-, \gamma$ ) [39, 40]

**Readout Electronics:** See Sec.8.7; CAMEX chip (CMOS):  $\sim 25$  krad powered ( $\sim 100$  krad unpowered) [41]; JAMEX chip (JFET):  $\sim 170$  krad powered ( $\sim 200$  krad unpowered) [42]; SVX chip:  $\sim 20$  krad [43]; Rad hard processes: see [44, 45, 46].

**Permanent Magnet Material:** SmCo,  $\sim 10^4$  Mrad, NdFeB,  $\sim 50$  Mrad ( $e^-, \gamma$ ),  $5 \times 10^{14}$  n/cm<sup>2</sup> [47, 48, 49]. See also Sec.6.4.

**Radiation budget** Whereas the detector occupancy rates are relevant during normal high-energy physics (HEP) data-taking, the detector radiation dose that determines the lifetime of components includes all sources. The total accumulated dose is the sum of HEP, injection, machine studies, accelerator startup (initial vacuum system processing), beam-loss events, etc. CESR Phase II measured the following: HEP (63%), injection (17%), machine studies (12%), other (8%) [5, 7].

#### 3.3.9.3 Detector background shielding

##### Synchrotron radiation shielding [20, 50, 51]

The detector is shielded from SR by the placement of a mask tip near the detector beam pipe. This mask tip is placed to intercept the SR fan from the nearby dipole which would ordinarily sweep directly across the detector beam pipe, as well as the radiation generated in IR quads. In high current machines, a thin metallic coating (usually high-Z) is applied to the beam pipe for further reduction of the SR flux entering the detector.

The placement and size of the innermost mask tip (or tips for asymmetric machines) determines the effectiveness of SR shielding, and the ultimate SR background rates that will be achieved, once the accelerator lattice is fixed. In addition to respecting the beam stay clear criteria, the following are general guidelines for choosing the mask tip radius and position considering SR shielding (a similar analysis is required for beam-gas):

1. The mask tip should be placed so that no direct radiation within  $\pm 10\sigma_x, \pm 10\sigma_y$  may strike the beam pipe.

2. Placement of the mask tip in  $s$  is a tradeoff between two factors: i) moving the tip closer to the IP decreases the portion of upstream vacuum chamber visible by the beam pipe, thereby protecting the detector beam pipe from scattered radiation upstream, and ii) tip scattering and backscattering rates from the mask increase rapidly as the mask tip is moved closer to the IP. These two concerns must be weighed, and a general guideline is to place the tip as close to the IP as acceptable background levels will allow.
3. The mask tip radial position should be made small enough to shield the beam pipe from direct SR, but not so small that it intercepts SR flux that would otherwise pass through the IR.
4. No surface within  $\sim 3$  m of the IP (for a  $B$ -factory) on which SR lands should be visible by the detector beam pipe (except the one irreducible backscattering surface). A consequence of this rule is that mask surfaces must either be sufficiently sloped so as to remain invisible to the central beam pipe, or smooth tapers must be stepped.
2. Collimators may be placed upstream of the IP to remove scattered beam particles. In cases where lost particle trajectories pass outside of the beam stay clear envelope it may be possible to significantly reduce the flux by placement of a collimator at points of high  $D, \beta$ . This technique is not possible for trajectories which remain within the beam-stay clear envelope, however. The PEP-II design calls for graded apertures, becoming progressively larger toward the IP [20]. Likewise, the DAΦNE design [23] calls for large tungsten beam scrapers upstream of each IR to be inserted to  $10\sigma_H$  reducing the flux of off-momentum particles striking the KLOE IR by a factor of  $\sim 200$ . LEP used a system of distributed adjustable tungsten collimators to remove off-energy beam particles in the arcs, and for detector shielding in the IR straights [8].
3. The detector beam pipe and vertex detector are shielded with nearby high- $Z$  masking material. The inner profile is designed by considering SR shielding requirements.

### 3.3.9.4 Detector background and radiation estimation

Detector background estimation is a three step process: (i) production of background generating particles, (ii) propagation of these particles to the detector region, and (iii) interaction and shower development in the detector and surrounding material. Generally, only the first step may be attacked analytically; the remainder of the problem is simulated with Monte-Carlo methods.

**Lost particle simulations** The general program proceeds as follows. Coulomb, bremsstrahlung, Touschek or Thermal Compton scattering events are generated along the beam orbit with a uniform scattering angle or energy loss probability. Scattered particle trajectories are tracked to the IP, and then trajectories striking within a few meters of the IP are collected and used as input to a full detector Monte-Carlo code including the solenoidal detector field and IR magnetic elements. The events are weighted by their production probabilities and the gas pressure at the point of scattering. The program DECAY TURTLE [55] can be used for generating and tracking of beam-gas scattered particles. Detector specific Monte-Carlo codes (generally GEANT-based [56]) are used for the charged particle/photon interactions and detector response.

### Sec.3.3: PARTICLE-MATTER INTERACTION

**SR simulations** The general program proceeds as follows. The beam is tracked through nearby magnetic elements and SR photons are generated tangent to the trajectory, with a spectrum determined by the local radius of curvature. Beamsize and  $x, x'$  correlation are taken into account [50]. SR photon trajectories are tracked to vacuum chamber surfaces and collected. These are then used as input to Monte-Carlo scattering codes for propagation of the x-rays from the vacuum chamber surfaces to the detector beam pipe and into the active volume of the detector. There is little uniformity in SR background calculation codes. A package of codes for generating (SRGEN) and scattering (SRSIM) are in use at CESR [50]. The scattering code is a single-scattering Monte-Carlo optimized for SR background calculations, and includes all K and L shell fluorescence physics. At PEP-II, QSRAD was used for photon generation, and the interaction was simulated with EGS [57]. Likewise, “home-grown” SR simulation codes were in use at LEP [8].

**Comparison with measurements** Since the background calculations are quite complex, comparison with measurements of detector backgrounds are extremely valuable. Iterations of calculation and measurement are necessary. Detailed comparisons of detector background simulation predictions with measurements have been undertaken at several labs. Generally good agreement was obtained at CESR/CLEO [4, 5, 7, 11], PEP-II/BaBar [9], KEK-B/Belle [10], and LEP [8].

### References

- [1] H. DeStaelber, AIP Proc. 214 (1990) 59
- [2] N. Toge, Proc. Int. Workshop on B-Factories: Accelerators and Experiments (1992) 44
- [3] S. Khan, B Factories, SLAC-400 (1992) 290
- [4] D. Cinabro, S. Henderson, H. Yamamoto, Int. Conf. High Energy Phys. (1994); D. Cinabro, 8th Mtg. of DPF (1994), World Scientific, p.1474
- [5] S. Henderson, PAC 97
- [6] Proc. 2nd Workshop on Backgrounds at the Machine-Detector Interface, World Sci. (1997)
- [7] S. Henderson, in Ref.[6]
- [8] G. von Holtey et al, CERN-SL/97-40 (EA)
- [9] B. Aubert et al, NIM A479 (2002) 1; T. Mattison et al, PAC 99; W. Kozanecki et al, NIM A446 (2000) 59; T. Fieguth et al, PAC 05; W.S. Lockman et al, PAC 05
- [10] K. Akai et al, NIM A499 (2003) 191; T. Abe, H. Yamamoto, PRST-AB 7, 072802 (2004); J. Haba, e<sup>+</sup>e<sup>-</sup> Factories 1999, KEK Proc. 99-24, p.237
- [11] S. Henderson, D. Cinabro, in [10]
- [12] C. Zhang, Proc. of the 40th ICFA ABDW 2008; BEPCII group, BEPCII Design Report, IHEP Proc. 2001
- [13] M. Biagini et al, arXiv:1009.6178v1 (2010)
- [14] S. Hashimoto et al, KEK-REPORT-2004-4
- [15] J.H. Hubbell, I. Øverbo, J.Phys.Chem.Ref.Data 8 (1979) 69
- [16] J.H. Hubbell et al, J.Phys.Chem.Ref.Data 4 (1975) 471
- [17] E. Storm, H.I Israel, LA-3753 (1967)
- [18] M.O. Krause, J.Phys.Chem.Ref.Data 8 (1979) 307
- [19] D. Kirkby, in Ref.[6]
- [20] Asymmetric B-factory Conceptual Design Report, SLAC-372
- [21] M. Sullivan, PAC97 (1997) 296
- [22] V. Telnov, arXiv: 1203.6563 [hep-ex]
- [23] S. Guiducci, in Ref.[6]
- [24] R.A. Briere et al, Cornell CLNS 01/1742
- [25] B. Dehnig et al, PL B2349 (1990) 145
- [26] V.I. Telnov, NIM A260 (1987) 304
- [27] J. Turner et al, EPAC 04 Lucerne
- [28] Y. Funakoshi et al, EPAC 04
- [29] H. Ohkuma, EPAC 08; J. Corbett et al, IPAC 10; C. Cristou et al, IPAC 10
- [30] U. Wienands et al, EPAC 08
- [31] G. Bregliozi, Report at LHC Background Study Group, 18 Oct. 2010; cern.ch/lbs
- [32] C.L. Woody et al, IEEE TNS (1992) 524
- [33] Z. Wei, R. Zhu, NIM A326 (1993) 508
- [34] G. Bratta, H. Newman, R.Y. Zhu, Ann.Rev.Nucl.Part.Sci. 1994, 14:453
- [35] S. Majewski, C. Zorn, Instrumentation in High Energy Physics, World Scientific (1992) p.157
- [36] J. Kadyk, NIM A300 (1991) 436
- [37] J. Kadyk et al, IEEE TNS 37 (1999) 478
- [38] J. Va'vra, SLAC-PUB-5207
- [39] E. Fretwurst et al, NIM A288 (1990) 1-12
- [40] G. Hall, NIM A368 (1995) 199
- [41] J. Alexander et al, NIM A337 (1993) 171
- [42] C. Ward, Cornell CBX-96-71
- [43] N. Bacchetta et al, NIM A324 (1993) 284
- [44] L. Blanquart et al, IEEE TNS 41 (1994)
- [45] M. Raymond et al, NIM A351 (1994) 449
- [46] M. Millmore et al, 2nd Workshop on Electronics for LHC Experiments (1996) p.23
- [47] H.B. Luna et al, NIM A285 (1989) 349
- [48] W.V. Hassenzahl et al, NIM A291 (1990) 378
- [49] J. Pfluger, G. Heintze, I. Vasserman, RSI 66 (1995) 1946
- [50] CESR-B Conceptual Design for a B-factory Based on CESR (1993)

- [51] KEK-B Conceptual Design Report
- [52] S. Henderson, 8th Mtg. of DPF (1994) World Scientific, p.1480
- [53] S. Henderson, XXVI ICHEP (1992) 2022
- [54] D.J. Dumas et al, NIM A404 (1998) 17
- [55] D. Carey, K.L. Brown, Ch. Iselin, SLAC-246/UC-28/Fermilab PM-31 (1982)
- [56] R. Brun et al, CERN DD/EE/84-1 (1987)
- [57] W.R. Nelson, H. Hirayama, D.W.O. Rogers, SLAC-265 (1985)

### 3.3.10 Particle Interactions and Beam-Induced Backgrounds and Radiation

*N.V. Mokhov, S.I. Striganov, FNAL*

#### Hadronic and Electromagnetic Showers

Electromagnetic (EM) interactions, decays of unstable particles and strong inelastic and elastic nuclear interactions all affect the passage of high-energy particles through matter [1, 2, 3, 4]. At  $E >$  a few GeV the main feature of interactions is multi-particle production. Energetic particles are concentrated around the primary beam axis forming the shower core. Neutral particles (mainly neutrons) and photons dominate with a cascade development when  $E < 100$  MeV. The shower dimensions grow with primary energy. The nuclear interaction rate is determined by hadron-nucleus cross-sections (Fig.2, Sec.3.3.2). The EM showers induced mainly by  $\pi^0$ -decays have different scale, but several common features with hadron cascades as listed in Tab.1.

Ionization energy loss Conventional approach to EM processes is based on a continuous energy loss approximation using the mean stopping power for charged particles in material [1, 2, 3, 5]. A detailed Monte-Carlo approach to EM interactions of heavy particles is described in [6]. It is used in modern Monte-Carlo codes [7], where collisions of charged particles and atoms with energy transfer  $\varepsilon$  greater than a cutoff  $\varepsilon_c$  are considered as discrete events involving production of  $\delta$ -electrons,  $e^+e^-$ -pair, bremsstrahlung [6]. The secondaries with  $E > \varepsilon_c$  are followed explicitly. Energy losses with  $\varepsilon < \varepsilon_c$  (so-called restricted losses) are considered as continuous. The redefined Vavilov parameters [8] for restricted loss distribution are

$$\xi = Bs, \quad B = 0.1536 \frac{Z}{A\beta^2}$$

$$\kappa_n = \xi/\varepsilon_c, \quad \beta_n^2 = \beta^2 \varepsilon_c/\varepsilon_{\max} \quad (1)$$

where  $Z$  and  $A$  are the absorber atomic and mass numbers,  $\beta c$  is the particle velocity,  $s$  the path length in g/cm<sup>2</sup> and  $\varepsilon_{\max}$  the maximum energy transferred in a single collision. The log-normal distribution fits well the Vavilov function for  $\kappa_n > 0.3$  [9], and its use simplifies simulation drastically.

To simulate  $\delta$ -electron production at any step, one calculates  $\varepsilon_G = \xi/0.3$  and  $\varepsilon_\delta = \min(\varepsilon_G, \varepsilon_c)$ . Then, the restricted energy loss with  $E < \varepsilon_\delta$  is sampled from log-normal distribution, and the number of  $\delta$ -electrons with  $E > \varepsilon_\delta$  is simulated using a Poisson distribution. The coordinates of  $\delta$ -electrons are calculated recursively. The electron energies are sampled from the Bhabha formulae. Total energy loss of a particle is the sum of the  $\delta$ -electron energies and of the restricted energy loss. Correlations of ionization energy loss and angular deflection due to Coulomb scattering need to be taken into account in some applications. The algorithm is described in [9].

Radiative energy loss of heavy particles For muons (pions) with  $E >$  a few hundred GeV and protons with  $E >$  a few tens TeV, the radiative mechanisms (bremsstrahlung and direct  $e^+e^-$  pair production) dominate over the ionization losses, especially in heavy materials. The approximation [6] to the exact Kel'ner formulae can be used for the pair production cross-section,

$$\frac{d\sigma_p}{d\nu} = b_p \frac{a(1+a)}{\nu(\nu+a)^2}, \quad b_p = \left( \frac{1}{E} \frac{dE}{dx} \right)_p \quad (2)$$

where  $(dE/dx)_p$  is the mean energy loss per unit length,  $\nu = (\text{pair energy})/(\text{particle energy})$ . With a few percent accuracy,

$$b_p \approx 1.689 \times 10^{-5} \left[ \frac{\text{cm}^2}{\text{g}} \right] \frac{m_e}{M} \frac{Z(Z+1)}{A}$$

$$\times \left[ b_1 \ln \left( \frac{b_3 Z^{-1/3}}{1 + 4b_3 Z^{-1/3} M/E} \right) - b_2 \right]$$

Here  $M$  is the incident particle mass, respectively;  $a$  is determined from the second moment of the cross-section,

$$a = \begin{cases} a_1 \cdot 10^{-3} \frac{m_\mu}{M}, & E \leq a_3 \\ \left[ a_1 + a_2 \ln \left( \frac{E}{a_3} \right) \right] \cdot 10^{-3} \frac{m_\mu}{M}, & E > a_3 \end{cases} \quad (3)$$

$a_i, b_i$  for some charged particles are given in Tab.2. See also Sec.3.3.12.

There are a number of different approaches for the calculation of the muon bremsstrahlung cross-section. These methods differ mainly in

### Sec.3.3: PARTICLE-MATTER INTERACTION

Table 1: Comparison of the properties of electromagnetic and hadron cascades.

[\*] See Sec.3.3.12 for  $X_0$  (radiation length), and  $\lambda_I$  (interaction length) for various substances. See Sec.3.3.1 and [1] for  $E_c$  (critical energy).  $E_{\min}$  is minimal energy of detected particles.

| Characteristic   | Electromagnetic cascade  | Hadron cascade  |
|--|--|---|
| Multiplication processes   | Production of $e^-e^+$ pairs and electron bremsstrahlung   | Inelastic hadron-nucleus collisions   |
| Mean free path before the interaction                                      | $(9/7)X_0$ for photons, [*]<br>$X_0/\ln(E_e/E_\gamma)$ for electrons   | Mean free path of a hadron before the inelastic interactions:<br>$\lambda_I \approx A/N_A \sigma_{\text{in}}$ [*]                                   |
| Mean inelasticity for secondaries  | $\langle K \rangle \sim 1$ for $\gamma$ ,<br>$\langle K \rangle \sim 0.07$ for $e^-$   | $\langle K \rangle \sim 0.5$  |
| Secondary particles  | Electrons and photons  | Principally $\pi$ mesons and nucleons   |
| Longitudinal development of the cascade                                    | Depends weakly on matter if the depth $t$ is expressed in units of radiation length, $x = t/X_0$ .   | Depends weakly on matter if the depth $t$ is expressed in units of the mean free path of a hadron before inelastic interaction, $x = t/\lambda_I$ . |
| Energy dependence of cascade maximum position                              | $\sim \ln(E_0/E_{\min})$   | $\sim \ln(E_0/E_{\min})$  |
| Characteristic length of the cascade decay                                 | $\Lambda = (3 - 4)X_0$ , does not depend on $E_0$  | $\Lambda = (1 - 2)\lambda_I$ , increases with increasing $E_0$  |
| Length of the effective region of development of the cascade               | $t_{\text{eff}} \simeq (10 - 30)X_0 \cdot \ln(E_0/E_{\min})$   | $t_{\text{eff}} \simeq (5 - 10)\lambda_I \cdot \ln(E_0/E_{\min})$   |
| Transverse development of the cascade                                      | Depends weakly on matter if the distance from the shower axis is expressed in Moliere units, $r_M = E_s X_0 / E_c$ with $E_s \equiv m_e c^2 \sqrt{4\pi/\alpha} \approx 21.1$ MeV | Depends weakly on matter if the distance from the cascade axis is expressed in $\text{g/cm}^2$  |
| Transverse dimension of the effective region of development of the cascade | $r_{\text{eff}} \simeq 2r_M$   | $r_{\text{eff}} \sim \lambda_I$   |

the treatment of screening corrections. A general expression [6] allows arbitrary nuclear and atomic form-factors to be applied. For a small energy transfer  $\varepsilon < \varepsilon_\gamma = 10^{-3}E$  at a muon energy  $E > 10$  GeV, the bremsstrahlung differential cross-section reaches the complete screening limit

$$\Sigma_\gamma(E, \varepsilon) = \frac{d_\gamma}{\varepsilon}, \text{ where } \begin{cases} d_\gamma = \frac{4b_\gamma}{3} \\ b_\gamma = (\frac{3}{E} \frac{dE}{dx})_\gamma \end{cases} \quad (4)$$

With this, the restricted energy loss distribution

can be approximated by

$$f_c^\gamma(\Delta_c, E, s) \simeq \frac{d_\gamma s}{\varepsilon_c^{d_\gamma s}} \frac{1}{\Delta_c^{1-d_\gamma s}} \quad (5)$$

The continuous energy loss  $\Delta_c$  at a step  $s$  is sampled from here providing  $\Delta_c < \varepsilon_c < \varepsilon_\gamma$ . The production of bremsstrahlung photons with energies  $> \varepsilon_c$  is considered as a discrete process.

Deep inelastic interactions Existing models of deep inelastic muon-nucleus scattering are

Table 2: Parameters of the approximation for pair production cross section.

|                    | Muon<br>$E > 20 \text{ GeV}$ | Pion<br>$E > 20 \text{ GeV}$ | Kaon<br>$E > 50 \text{ GeV}$ | Proton<br>$E > 90 \text{ GeV}$ |
|--------------------|------------------------------|------------------------------|------------------------------|--------------------------------|
| $a_1$              | 5.2                          | 5.5                          | 5.45                         | 5.4                            |
| $a_2$              | .13                          | .18                          | .35                          | .43                            |
| $a_3 (\text{GeV})$ | 200                          | 200                          | 100                          | 100                            |
| $b_1$              | .787                         | .791                         | .819                         | .833                           |
| $b_2$              | 1.1                          | 1.09                         | 1.14                         | 1.17                           |
| $b_3$              | 2986                         | 3017                         | 2773                         | 2532                           |

consistent at the 30% accuracy level (see, e.g., [10]). However, the relative mean energy loss for this process is  $\leq 10\%$  of the total, even at very high energies. The corresponding mean free path exceeds  $\sim 100 \text{ m}$  of iron for muons in the TeV energy region.

Hadron-nucleus interactions Elastic  $hA$  interactions are described in Sec.3.3.2.1. Inelastic  $hA$  interactions are described by either Monte Carlo event generators or phenomenological models [2]. Theoretical calculations based on the intranuclear cascade model are reliable at proton momenta  $p_0 < 5 \text{ GeV}/c$ . Microscopic models, such as DPMJET [11] (based on the dual topological unitarization approach) do a decent job at higher energies  $> 10 \text{ GeV}/c$ . The entire kinematical region for the inclusive spectra can be reliably described with phenomenological models such as [12].

Many reliable data and parameterizations exist on pion yield in pp-collisions. One can compensate for the lack of data for  $pA$  reactions by using the following form for the double differential cross section of the  $pA \rightarrow \pi^\pm X$  reaction:

$$\frac{d^2\sigma_{pA \rightarrow \pi^\pm X}}{dpd\Omega} = R^{pA \rightarrow \pi^\pm X}(A, p_0, p, p_\perp) \frac{d^2\sigma_{pp \rightarrow \pi^\pm X}}{dpd\Omega} \quad (6)$$

where  $p$  and  $p_\perp$  are the total and transverse momenta of  $\pi^\pm$ , and  $A$  is an atomic mass of the target nucleus. The function  $R^{pA \rightarrow \pi^\pm X}$ , measured with much higher precision than the absolute yields, is almost independent of  $p_\perp$  and its dependence on  $p_0$  and  $p$  is much weaker than for the differential cross-section itself. Because of rather different properties of pion production on nuclei in the forward ( $x_F > 0$ ) and backward ( $x_F < 0$ ) hemispheres, where  $x_F$  is the Feynman's longitudinal variable, one treats these two regions differently.

$R(x_F > 0.05)$  In this region one assumes  $R^{pA \rightarrow \pi^\pm X} \sim A^\alpha$ . The power  $\alpha$  is almost

independent of the pion charge. The following parameterization was proposed [13] for  $p_0 \geq 70 \text{ GeV}/c$  and  $p_\perp \leq 1 \text{ GeV}/c$ :

$$\alpha_g = 0.8 - 0.75 x_F + 0.45 x_F^3 / |x_F| + 0.1 p_\perp^2 \quad (7)$$

It turns out that Eq.(7) describes data well at  $p_0 \geq 24 \text{ GeV}/c$  and can be successfully used at lower momenta ( $5 \leq p_0 \leq 24 \text{ GeV}/c$ ) if it is replaced with

$$\alpha = \alpha_g - 0.0087(24 - p_0) \quad (8)$$

The  $R^{pA \rightarrow \pi^\pm X} \sim A^\alpha$  form does not extrapolate well to  $A = 1$  because of the difference in the  $\pi$ -yield in pp and pn collisions. This difference can be taken into account if one uses the following form for  $R^{pA \rightarrow \pi^\pm X}$  [13]:

$$R^{pA \rightarrow \pi^\pm X} = \left(\frac{A}{2}\right)^\alpha f(p_0, Y) \quad (9)$$

where  $f(p_0, Y) = \frac{d\sigma}{dp}(pd \rightarrow \pi^\pm) / \frac{d\sigma}{dp}(pp \rightarrow \pi^\pm)$ . It turns out that pion yields in pd and pp collisions are not very different, i.e.  $f(p_0, Y) \approx 1$ . One finds  $f(p_0, Y)_{\pi^-} = 1 + 0.225/N_{\pi^-} - a_{\pi^-} Y_{cms}$ , where  $N_{\pi^-}$  is mean  $\pi^-$  multiplicity in pp collisions and  $Y_{cms}$  is pion rapidity in the CM. Data show linear dependence of  $N_{\pi^-}$  on free energy  $W = (\sqrt{s} - 2 \cdot m_p)^{0.75} / \sqrt{s}^{0.25}$ , where  $\sqrt{s}$  is the CM collision energy. Our fit to the data gives  $N_{\pi^-} = 0.81(W - 0.6)$ . The other parameter  $a_{\pi^-} = 0.16$  for  $p_0 \leq 20 \text{ GeV}/c$ , and depends on energy for higher momenta as  $a_{\pi^-} = -0.055 + 0.747/\ln(s)$ .  $f(p_0, Y)_{\pi^+}$  is forced to be 1 if it becomes less than 1. For  $\pi^+$  production the approximation is much simpler  $f(p_0, Y)_{\pi^+} = 0.85 + 0.005 p_0$  for  $p_0 \leq 30 \text{ GeV}/c$  and  $f(p_0, Y)_{\pi^+} = 1$  for higher momenta.

$R(x_F < 0.05)$  In this region, due to the lack of experimental data on  $\alpha$ , one uses the following expression for the function  $R$  in Eq.(6):

$$R^{pA \rightarrow \pi^\pm X} = \frac{dN/dY(pA)}{dN/dY(pp)} \quad (10)$$

### Sec.3.3: PARTICLE-MATTER INTERACTION

The following scaling law was proposed in [14] for charged shower particle ( $\beta > 0.7$ ) production in  $pA$  collisions at 20–400 GeV/ $c$ :

$$\frac{Y_0}{\langle N_s \rangle} \frac{dN}{d\eta} = f \left( A, \frac{\eta}{Y_0} \right) \quad (11)$$

where  $\langle N_s \rangle$  is a mean multiplicity of shower particles,  $Y_0$  is rapidity of primary proton and  $\eta = -\ln(\tan(\theta/2))$  is pseudorapidity of a secondary particle. One finds that this approximation is in a reasonable agreement with data at  $p_0 > 7.5$  GeV/ $c$ . Unfortunately,  $\eta$  is not a convenient variable to describe forward pion production ( $\theta \approx 0$ ). Analysis [12] of the  $pA \rightarrow \pi^- X$  data at  $10 < p_0 < 100$  GeV/ $c$  shows that replacing  $\eta$  in Eq.(11) with rapidity  $Y$

$$\frac{dN}{dY} = \frac{\langle N_\pi \rangle}{Y_0} F \left( A, \frac{Y}{Y_0} \right) \quad (12)$$

provides better description of the pion yield in the entire kinematic range. Here  $\langle N_\pi \rangle$  is mean pion multiplicity,  $Y_0 = \ln(\frac{E_0 + p_0}{m_p})$  is rapidity of incident proton and  $Y = \ln(\frac{E_\pi + p_z}{m_\pi})$  is  $\pi$  rapidity,  $m_\perp = \sqrt{p_\perp^2 + m_\pi^2}$ . One chooses the Gaussian form for the scaling function

$$F \left( A, \frac{Y}{Y_0} \right) = c_1 \exp \left[ - \left( \frac{Y}{Y_0} - c_2 \right)^2 / c_3 \right] \quad (13)$$

where for  $\pi^-$ :  $c_1 = 1.149 A^{0.0479}$ ,  $c_2 = 0.492 A^{-0.0565}$ , and  $c_3 = 0.214 A^{-0.121}$ . Reliable rapidity distributions for  $\pi^+$  at  $x_F < 0$  are measured only for  $p_0 \geq 100$  GeV/ $c$ . Assuming that the scaling (15)–(16) is valid for  $\pi^+$  also, one finds the following parameters from data:  $c_1 = 1.6$ ,  $c_2 = 0.4833 - 0.0006A$ ,  $A < 108$  and  $c_2 = 0.418$ ,  $A > 108$ ,  $c_3 = 0.1873 - 0.0027A$ ,  $A < 24.3$  and  $c_3 = 0.122$ ,  $A > 24.3$ . The data on  $dN/dY$  for  $pp \rightarrow \pi^\pm X$  reaction is well described by:  $dN/dY = C_{pp} \exp(-Y_{cm}^2/2\sigma^2 + dY_{cm}^4)$ , where  $\sigma_{\pi^+} = -2.725 + 2.065 \ln(\sqrt{s}) + 4.205/\sqrt{s}$  and  $\sigma_{\pi^-} = -0.732 + 1.1 \ln(\sqrt{s}) + 1.3/\sqrt{s}$ . The parameter  $d$  depends on energy as  $d = w/s^\gamma - 1/R$ , where  $w_{\pi^+} = 6.68$ ,  $\gamma_{\pi^+} = 2.1$ ,  $w_{\pi^-} = 496$  and  $\gamma_{\pi^-} = 4$ . Coefficient  $R$  reads as  $R_{\pi^+} = 41.5 + 0.151(\ln(s))^{3.29}$  and  $R_{\pi^-} = 63.2 + 0.052(\ln(s))^{3.77}$ . The normalization parameter in (13), combined of  $\langle N_\pi \rangle$ ,  $C_{pp}$  etc, is chosen to match the functions (12) and (13) at  $x_F = 0.05$ .

$pp \rightarrow \pi^\pm X$  To describe the invariant cross section of charged pion production in  $pp$

Table 3: Parameters in Eq.(14).

|         | A    | B   | C    | D   | E  | F   |
|---------|------|-----|------|-----|----|-----|
| $\pi^+$ | 60.1 | 1.9 | 0.18 | 0.3 | 12 | 2.7 |
| $\pi^-$ | 51.2 | 2.6 | 0.17 | 0.3 | 12 | 2.7 |

collisions one can use the formula

$$E \frac{d^3 \sigma^{pp \rightarrow \pi^\pm X}}{dp^3} = A \left( 1 - \frac{p^*}{p_{\max}^*} \right)^B \times \exp \left( -\frac{p^*}{C\sqrt{s}} \right) V_1(p_\perp) V_2(p_\perp) \quad (14)$$

where  $p^*$  and  $p_{\max}^*$  are pion momentum and maximum momentum transfer in CM and parameters are given in Tab.3. The best description of the  $p_\perp$  dependence is obtained with:

$$V_1(p_\perp) = (1 - D)e^{-Ep_\perp^2} + De^{-Fp_\perp^2}, \\ p_\perp \leq 0.933 \text{ GeV}/c$$

$$V_1(p_\perp) = 0.2625/(p_\perp^2 + 0.87)^4, \\ p_\perp > 0.933 \text{ GeV}/c \quad (15)$$

$$V_2(p_\perp) = 0.7363e^{0.875p_\perp}, \\ p_\perp \leq 0.35 \text{ GeV}/c \quad (16)$$

$$V_2(p_\perp) = 1, p_\perp > 0.35 \text{ GeV}/c$$

$\gamma A \rightarrow hX$  reactions Hadroproduction in photon-nucleus interactions at  $E_\gamma \geq 0.14$  GeV can be simulated approximately by replacing the photon with a real pion of random charge with the same kinetic energy. The total cross-section is calculated as  $\sigma_{\gamma A} = R_A [Z\sigma_{\gamma p} + (A - Z)\sigma_{\gamma n}]$ . For the total  $\gamma p$  cross-section experimental data are used [6]. The  $A$ -dependence of the cross-section  $R_A$  is extracted from experimental data, with  $R_A \approx 1.047 A^{-0.085}$  at high energies. A comparison of this description with data is shown in Fig.1. Photoneutron production in the giant resonant energy region  $6 \leq E_\gamma \leq 60$  MeV is described according to the algorithm [15, 16] extended for light nuclei  $4 \leq A \leq 56$  on the basis of the latest data. An interpolation is used between 60 and 140 MeV.

**Space-time source term** A detailed comparison of detector backgrounds for hadron, electron and muon colliders is given in [17]. LHC, NLC (or ILC nowadays) and  $2 \times 2$  TeV  $\mu^+ \mu^-$  colliders are considered representatives of their classes.

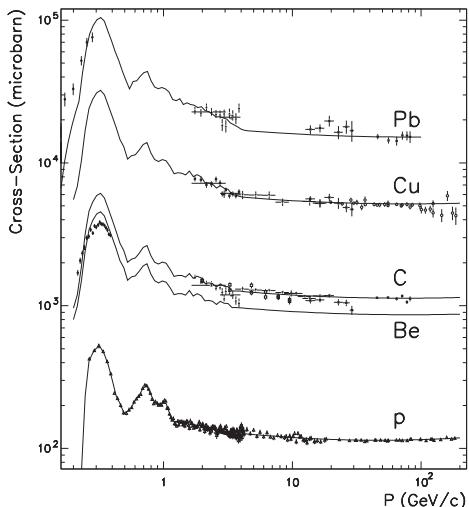


Figure 1: Calculated photon-nucleus cross section in comparison with experimental data vs photon momentum for lead, copper, carbon, beryllium and hydrogen.

Despite different colliding particle types and machine parameters, there are common considerations for these accelerators.

Collision remnants from the IP are often the major source of background and radiation levels in hadron collider detectors, in experimental halls and in final focus quads. Small aperture collimators on either side of the IP in front of the first final focus quads protect the accelerator components from IP radiation. Shielding around collimators reduces radiation levels in detectors and in collision halls.

Beam loss in the IP vicinity is the second background source. Without protection, the number of hits from halo particles in the detector can be greater than that from the IP and their products. Reduction of beam loss in the IR improves this situation. Multi-stage high-performance collimation system (Sec.3.3.11) is the major way to do that. Tertiary collimators at the non-IP sides of the IR with additional collimators as close as a few centimeters from the IP (at lepton colliders) are mandatory nowadays. Plugging the accelerator tunnel at the entrance to the experimental hall further helps reduce this component.

Temporal considerations in the background analysis are important. Integrated levels determine radiation damage, aging and radioactivation

of detector components, and the radiation environment in the experimental hall. High instantaneous particle fluxes complicate track reconstruction, cause increased trigger rates, and affect detector occupancy. Therefore, the beam's time structure must be taken into account in the comparison of expected background environments at different machines.

Instantaneous and integrated rates Parameters of three types of future colliders are presented in Tab.4. The *integrated* luminosities are obtained from the design peak values by multiplying by  $10^7$  s. The *instantaneous* or *effective* luminosity, which determines the detector performance, is defined for the amount of radiation in the detector active element over the drifting time  $\Delta t_d$  or the bunch train length, whichever is smaller. Collider detector elements most susceptible to occupancy problems have the drifting/integration time in the 40 to 300 ns range. Taking conservatively  $\Delta t_d = 300$  ns, one gets  $\mathcal{L}_{\text{eff}} = \mathcal{L} \times \Delta t_d$  for LHC and  $\mathcal{L}_{\text{eff}} = \mathcal{L}/f$  for the other colliders, where  $f$  is the repetition rate (collision rate for  $\mu^+ \mu^-$ ). The NLC (ILC)  $\gamma\gamma$  mode is 10 to 20 times worse than the  $e^+ e^-$  one because of the large backscattered laser cross-section [18]. Multiplying the above luminosities by the appropriate hadron production cross-sections, one can compare pp,  $e^+ e^-$  and  $\mu^+ \mu^-$  collisions as a source of background. The last two lines of Tab.4 show that LHC produces at least  $10^6$  times more background hadrons from the IP annually than the lepton machines. At the same time the instantaneous background productions are not so drastically different.

The situation is very different with the accelerator related backgrounds. This component is due to elastic and inelastic beam-gas interactions in the collider lattice, and due to quasi-local beam halo loss in the IR components, mainly diffractive protons from another IP, and halo tails from the beam cleaning system. At the high-energy lepton colliders the backgrounds generated in the machine are a major concern. Synchrotron radiation and muons produced in beam halo interactions along the lattice create serious backgrounds in the detectors at linear  $e^+ e^-$  colliders. These can be reduced with an appropriate final focus design and a set of collimators (Secs.3.3.9,3.3.11). At high energy muon colliders the situation is much worse. Unavoidable  $\mu \rightarrow e\nu\bar{\nu}$  decays occurring in the beam-pipe are the major source [6, 17, 19, 20, 21, 22].

### Sec.3.3: PARTICLE-MATTER INTERACTION

Table 4: Collider parameters and calculated integrated and *effective* luminosities.

| Parameters  | LHC                   | NLC-500               | NLC-1000              | $\mu^+\mu^-$          |
|---|-----------------------|-----------------------|-----------------------|-----------------------|
| $E_{cm}(\text{TeV})$  | 14                    | 0.5                   | 1                     | 4                     |
| $\mathcal{L}(10^{34}\text{cm}^{-2}\text{s}^{-1})$   | 1                     | 0.71                  | 1.45                  | 4.55                  |
| Rep. (collision) rate $f$ (Hz)  | -                     | 180                   | 120                   | $4.04 \times 10^4$    |
| Particles/bunch ( $10^{11}$ )   | 1                     | 0.07                  | 0.11                  | 20                    |
| Bunch/pulse   | -                     | 90                    | 75                    | 1                     |
| Bunch separation (ns)   | 25                    | 1.4                   | 1.4                   | $18.6 \times 10^3$    |
| Yearly $\mathcal{L}_y(\text{fb}^{-1})$  | 100                   | 71                    | 145                   | 455                   |
| $\sigma_h(\mu\text{b})$   | $80 \times 10^3$      | 0.045                 | 0.034                 | 0.054                 |
| $\Delta t_d$ or bunch train length (ns)   | 300                   | 126                   | 105                   | -                     |
| $\mathcal{L}_{\text{eff}}(\text{cm}^{-2})$  | $3.00 \times 10^{27}$ | $3.94 \times 10^{31}$ | $1.21 \times 10^{32}$ | $1.13 \times 10^{30}$ |
| $(\sigma_h \times \mathcal{L}_y)/(\sigma_h \times \mathcal{L}_y)_{LHC}$                       | 1                     | $4.00 \times 10^{-7}$ | $6.16 \times 10^{-7}$ | $3.07 \times 10^{-7}$ |
| $(\sigma_h \times \mathcal{L}_{\text{eff}})/(\sigma_h \times \mathcal{L}_{\text{eff}})_{LHC}$ | 1                     | $7.39 \times 10^{-3}$ | $1.71 \times 10^{-2}$ | $2.54 \times 10^{-4}$ |

#### Hadron colliders

Interaction point The charged particle fluxes in the detector tracking cavity, determined by the primary events, decrease as  $1/r^2$ , whereas neutron fluxes are more uniform in the cavity, depending strongly on the calorimeter material. The damage induced in semiconductor components is linearly dependent on the non-ionizing energy loss (NIEL  $\sim 100$  MeV mb), determined by the integrated flux of  $> 100$  keV neutrons and charged hadrons. The hit rates are proportional to the charged particle flux (primary and that created by neutrals) in the sensitive volume and are related to the effective luminosity. The rates in the endcap calorimeters are much higher, especially at small radii. Neutron fluxes in the electromagnetic calorimeter can reach high values severely restricting the lifetime of silicon detectors and readout electronics. In the forward muon system, the signal is composed of charged particles, photons with  $\sim 1\%$  efficiency and neutrons ( $\sim 0.3\%$  efficiency). For the CMS and ATLAS detectors, the signal rate ranges from  $600 \text{ Hz/cm}^2$  at  $r = 1 \text{ m}$  to a few  $\text{Hz/cm}^2$  at  $r > 5 \text{ m}$ .

Beam halo The crucial issue here is the beam loss distribution in the IR [23, 24]. The first component is calculated via modeling beam interactions with residual gas in the beam pipe using a detailed gas pressure map for the entire machine [25]. For the LHC, the nuclear inelastic interaction rates are approximately  $500 \text{ m}^{-1}\text{s}^{-1}$  in warm and cold straights and  $2 \times 10^4 \text{ m}^{-1}\text{s}^{-1}$  in cold arcs. The second component is formed by interactions with the IR limiting apertures of a tertiary halo of protons escaping the betatron and momentum cleaning collimation systems.

The mean energies of particles coming to the detector from the LHC tunnel are [23]  $6.6 \text{ GeV}$  ( $\mu$ ),  $8.1 \text{ GeV}$  ( $h^\pm$ ),  $310 \text{ MeV}$  (n),  $150 \text{ MeV}$  (e $^\pm$ ), and  $30 \text{ MeV}$  ( $\gamma$ ). The mean distance from the beam axis is  $1.5 \text{ m}$ , the mean angles are  $\sim 600 \text{ mrad}$  for neutrons and  $\sim 130 \text{ mrad}$  for all other particles. Most of these particles can be intercepted with a concrete plug at the tunnel-hall interface, except for muons, which penetrate through shielding, accelerator and detector, creating a rate of a few  $\text{Hz/cm}^2$  in the detector.

#### e $^+$ e $^-$ linear colliders

Interaction point Tab.5 shows that the average integrated hadronic fluxes produced at the IP at a linear collider (ILC project) are about  $10^6$  lower compared to LHC. However, the difference in instantaneous rates is less drastic.

Beam halo Synchrotron radiation, beam-gas and beam halo interactions with the components of the final focus and adjacent sections of the linear colliders create fluxes of muons and other secondaries which can exceed the tolerable levels at a detector by a few orders of magnitude. A multi-stage collimation set and a system of magnetized iron spoilers which fill the tunnel can meet the ILC design goal of allowing a continuous  $0.1\%$  beam loss, resulting in a few muons at the detector [26, 27].

#### $\mu^+\mu^-$ colliders

Interaction point A muon collider is the “cleanest” machine with respect to both integrated and instantaneous particle background from the IP (see Tab.5).

Beam decays One of the most challenging problems in the design of a muon collider arises from muon decay [6, 17, 20, 22]. With  $2 \times 10^{12}$

Table 5: Background fluxes ( $\text{cm}^{-2}$ ) from the IP accumulated over 1 year (1) and *effective* (2) in central tracker, endcap calorimeter and forward muon spectrometer at different radii.

| Detector | r (cm) | LHC               | ILC            | $\mu^+\mu^-$      |
|----------|--------|-------------------|----------------|-------------------|
| (1)      |        |                   |                |                   |
| Tracker  | 30     | $2\times 10^{13}$ | $10^7$         | $6\times 10^6$    |
| ECAL     | 50     | $10^{14}$         | $10^8$         | $10^8$            |
| Forward  | 100    | $10^{11}$         | $5\times 10^3$ | $8\times 10^3$    |
| (2)      |        |                   |                |                   |
| Tracker  | 30     | 0.6               | 0.01           | $2\times 10^{-4}$ |
| ECAL     | 50     | 0.9               | 0.8            | $2\times 10^{-2}$ |

Table 6: Accumulated over 1 year and *effective* accelerator related fluxes ( $\text{cm}^{-2}$ ) in detector components at  $r = 50 \text{ cm}$ , with all the protective measures on.

|            | LHC               | ILC-1000         | $\mu^+\mu^-$ |
|------------|-------------------|------------------|--------------|
| Integrated | $10^8$            | $1.6\times 10^6$ | $10^{14}$    |
| Effective  | $3\times 10^{-6}$ | $10^{-3}$        | 5            |

muons in a 2-TeV bunch, there are  $2\times 10^5 \mu\rightarrow e\nu\bar{\nu}$  decays per meter in a single pass through an IR, or  $6\times 10^9$  decays/m/s. Both the decay electrons ( $E \approx 700 \text{ GeV}$ ) and the synchrotron photons emitted by these electrons in a strong magnetic field induce EM showers in the collider and detector components. Almost 15 MW of power is deposited in the storage ring, or about 2 kW/m. The resulting heat load to the cryogenic systems and the background levels in the collider detectors are serious issues. The intense, directed neutrino beam generated is perhaps an issue more serious than the high radiation level from EM showers. Neutrino interactions in the soil intended to shield the collider may produce unacceptable radiation levels at large distances.

By carefully designing the final focus system with special spoiling dipoles, by embedding sophisticated collimators in the immediate IP vicinity and shielding and other protective measures the heat load and backgrounds can in principle be mitigated by several orders of magnitude [17, 20]. All particles over a wide energy range contribute to the background levels. Mean momenta of particles entering a 1.5-TeV muon collider detector are[22]: 0.9 MeV/c ( $\gamma$ ), 6 MeV/c ( $e^\pm$ ), 45 MeV/c (n) 0.5 GeV/c (charged hadrons) and 23 GeV/c

( $\mu$ ). There is a rather uniform distribution of neutrals in the cavity with charged fluxes being substantially lower. The maximum hit rate density in a vertex detector of a 1.5-TeV muon collider is calculated to be about  $1 \text{ MHz/mm}^2$ , same as in the CMS detector at the LHC at the same luminosity. At the same time, total the peak hit rates per bunch crossing at the muon collider are higher than at the LHC. It is assumed that a reliable beam cleaning system is in the lattice far upstream from the IP. Studies show that the loss of even a small fraction of the beam within a few hundred meters of the IP results in backgrounds in a detector comparable to those from  $\mu\rightarrow e\nu\bar{\nu}$  decays.

Background particle spectra and space distributions are not very different in similar detector configurations at hadron,  $e^+e^-$  and  $\mu^+\mu^-$  colliders. Expected background levels are summarized in Tab.5 for IP and in Tab.6 for accelerator backgrounds. The *integrated* fluxes are converted to the NIEL or “equivalent 1 MeV neutron” values, and *instantaneous (effective)* fluxes (signals) are defined as the charged particle flux plus  $0.003 \times (F_n + F_\gamma)$ .

## References

- [1] Review of Particle Physics, PLB 592 (2004)
- [2] A.N. Kalinovskii, N.V. Mokhov, Yu.P. Nikitin, Passage of High-Energy Particles through Matter, AIP (1989)
- [3] J.D. Jackson, Classical Electrodynamics, 3rd ed., Wiley (1999)
- [4] A. Ferrari, P. Sala, The Physics of High Energy Reactions, Int. Centre for Theoretical Physics (1996)
- [5] D.E. Groom, N.V. Mokhov, S.I. Striganov, Atomic Data and Nuclear Data Tables, 78 (2001) 183
- [6] N.V. Mokhov, S.I. Striganov, AIP Proc. 372 (1995) 234
- [7] N.V. Mokhov, FNAL-FN-628 (1995); Fermilab-Conf-04/053 (2004); <http://www-ap.fnal.gov/MARS/>
- [8] P.V. Vavilov, JETP 5 (1957) 749
- [9] S.I. Striganov, Rad. Prot. Dosimetry 116 (2005) 293-296
- [10] L.B. Bezrukov, E.V. Bugaev, Sov. J. Nucl. Phys. 33 (1981) 635
- [11] S. Roesler, R. Engel, J. Ranft, ICRC-2001, Copernicus Gesellschaft (2001) 439
- [12] N.V. Mokhov, S.I. Striganov, AIP Proc. 435 (1997) 453
- [13] W.M. Geist, Nucl. Phys. A525 (1991) 149

### Sec.3.3: PARTICLE-MATTER INTERACTION

- [14] E. Stenlund, I. Ottlund, CERN-EP/82-42 (1982)
- [15] I.S. Baishev, I.A. Kurochkin, N.V. Mokhov, IHEP-91-118, Protvino (1991)
- [16] B.L. Berman, S.C. Fultz, RMP 47 (1975) 713
- [17] N.V. Mokhov, Nucl. Phys. B (Proc. Suppl.) 51A (1996) 210
- [18] R. Engel et al, Proc. 2nd Workshop on Simulating Accelerator Radiation Environments (SARE2) (1995)
- [19]  $\mu^+ \mu^-$  Collider Feasibility Study, BNL-52503; Fermilab-Conf-96/092; LBNL-38946 (1996)
- [20] C. Johnstone, N. Mokhov, Fermilab-Conf-96/366 (1996)
- [21] N. Mokhov et al, PAC 11 (2011); also arXiv:1202.3979
- [22] N. Mokhov S. Striganov, Physics Procedia (2011), arXiv:1204.6721
- [23] A. Drozhdin, M. Huhtinen, N. Mokhov, NIM A381 (1996) 531
- [24] N. Mokhov, T. Weiler, CERN-2009-003 (2009) 37
- [25] A. Drozhdin, N. Mokhov, S. Striganov, PAC 09, Fermilab-Conf-09/172-APC
- [26] A. Drozhdin et al, Fermilab-TM-2200 (2003)
- [27] N.V. Mokhov, A.I. Drozhdin, M.A. Kostin, PAC 05

#### 3.3.11 Beam Collimation

*R. Assmann, CERN*

**Stored energy and density of beams** The beam current  $i_b$  is maximized by increasing the repetition frequency (or revolution frequency)  $f_r$  of the accelerator and by increasing the number of particles  $N_p = i_b/(f_r q)$ , with  $q$  being the charge of a beam particle. Considering single charge particles ( $q = e$ ) with relativistic momentum  $p$  ( $\gamma \gg 1$ ), a beam with  $N_p$  particles carries a stored energy  $E_{\text{stored}} = p c e N_p$ . Here,  $c$  is the light velocity. Fig.1 compares the stored energy for various electron and proton accelerators [1]. Modern accelerators operate with beams between 10 kJ and 500 MJ. Most of the listed accelerators are storage rings where the beam is exploited for many hours. However, linac-based facilities (SNS, ILC, CLIC) use the produced beam only once and regenerate it at 5–100 Hz. Loss-induced heating can be in the range from 1 kW to 500 kW if 0.1% of beam is lost over 1 s.

Most accelerators require a high brightness, achieved by small transverse beam sizes  $\sigma_x$  and  $\sigma_y$ . As a consequence the energy density  $\rho_e =$

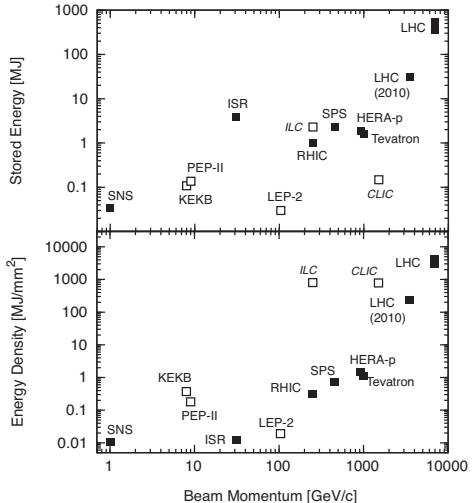


Figure 1: Stored energy (top) and stored energy density (bottom) for different electron (open symbol) and proton (filled symbol) accelerators. ILC and CLIC are design studies for  $e^+e^-$  colliders. Stored energy density is calculated at locations of collimators or other targets.

$E_{\text{stored}}/(2\pi\sigma_x\sigma_y)$  is increased. Energy densities for various accelerators are shown in Fig.1 [1]. Modern accelerators operate with beam densities in the range of  $10 \text{ kJ/mm}^2$  to  $4 \text{ GJ/mm}^2$ . This is often far above the typical damage limit of metals ( $50 \text{ kJ/mm}^2$ ).

**Collimator concepts** Collimators place special blocks of materials (“jaws”) at strategic locations such that beam losses are safely absorbed, protecting against damage and loss-induced perturbations of the accelerator operation. The collimator jaws are the materials which are closest to the beam and usually must be especially robust. Many different concepts exist. A few different collimator schemes are illustrated in Fig.2. Collimators can be as simple as a fixed vacuum pipe with thick walls, sometimes movable as a whole. At the other extreme they can be advanced high tech devices which consist of a vacuum tank, two fully movable jaws, cooling, stepping motors and redundant position diagnostics. This design type (as developed for the LHC) is illustrated in Fig.3.

Collimator jaws are either slightly curved along the beam direction (to reduce sensitivity to beam-jaw angular alignment) or have a fully flat top. Magnetized jaws have been considered and

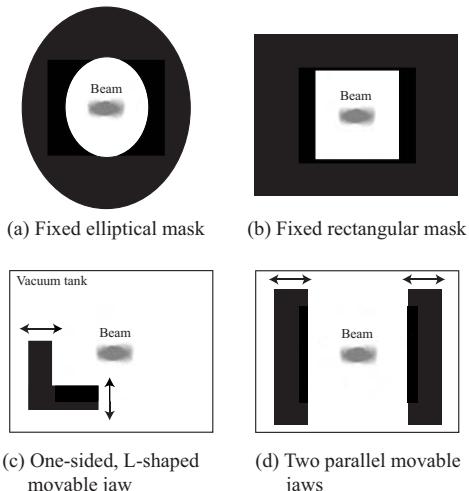


Figure 2: Illustration of possible collimator concepts. All designs can be installed in different azimuthal angles to provide collimation in horizontal, vertical and/or skewed planes.

might have some potential for very high beam energies. So far, however, magnetized jaws have not been used.

**System design goals** Several strategically placed collimators form a collimation system that has to fulfill various critical tasks [1, 2, 3]:

(i) Background control: Beam halo, beam-gas scattering and emitted synchrotron photons can induce spurious signals in physics detectors. Materials with good absorption quality are placed around the experiment and shield the sensitive detectors.

(ii) Hands-on maintenance: Collimators intercept beam losses at locations optimized for radiation impact. A typical design goal is to keep beam loss outside of collimators below the 1 W/m level allowing hands-on maintenance (residual dose rate  $< 1 \text{ mSv/h}$  at 30 cm). Collimators must have good absorption quality and are often heavily shielded.

(iii) Protection against heating: Direct beam loss or synchrotron radiation losses can lead to local heating. Typical quench limits for superconducting magnets range from  $1 \text{ mW/cm}^3$  to  $100 \text{ mW/cm}^3$ , to be compared with possible power loss in the kW to MW regime. Beam loss related heating is minimized by highly efficient multi-stage collimation sections ( $\text{leakage} \ll 0.1\%$ ).

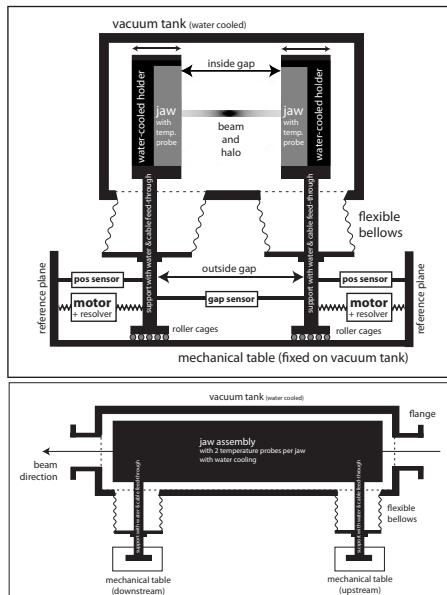


Figure 3: Sketch of an advanced collimator design with two movable jaws inside a vacuum tank (top: view along beam direction; bottom: side view). The jaws are supported by two mechanical tables, one at each end of the tank. Reference planes allow precise calibration during production and outside monitoring of absolute jaw positions and gaps.

Heating from synchrotron-radiation is intercepted with local collimators.

(iv) Ultra-high vacuum: Beam loss induced desorption can reduce beam lifetimes. This is particularly important for partially ionized beams: ions lose an electron by beam-gas scattering, affected ions get lost in a high dispersion point, desorbed molecules degrade the vacuum, more ions experience beam-gas events, ... and a vacuum instability develops. Collimators with low desorption materials are designed to intercept losses instead.

(v) Passive protection: High power beams can potentially destroy parts of the accelerator. Collimators are placed at strategic locations to intercept mis-directed beam, for example from kicker errors (timing, amplitude) and trips of accelerator equipment (dipoles, quadrupoles, correctors, rf). Exposed materials must be robust. Additional collimators are often placed behind the exposed collimators for absorption of diluted beam particles and showers.

### Sec.3.3: PARTICLE-MATTER INTERACTION

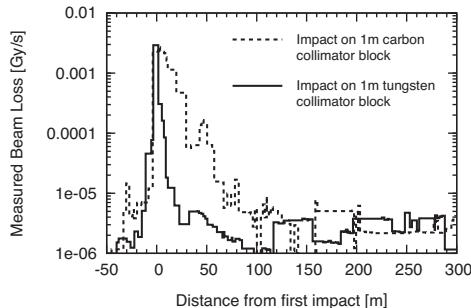


Figure 4: Measured beam loss profile (BLM) downstream of a single collimator for two different materials when hit by 450 GeV protons.

**Effect of a single collimator** A collimator jaw intercepts stray particles, but can also give rise to a wakefield (or impedance).

When a particle hits a collimator jaw then the particle experiences a number of elastic and inelastic processes (Secs.3.3.1,3.3.10). An inelastic interaction initiates a downstream shower and the initial particle is usually lost locally (often called “absorbed”). The probability for inelastic interaction and the amount of energy absorbed in the collimator block depend on the type of beam, its energy, the jaw material used, the jaw length and transverse dimensions [4]. They are calculated with dedicated specialist tools like FLUKA, MARS and GEANT. Measurements are shown in Fig.4 for the LHC beam, comparing measured beam loss for 450 GeV protons impacting on a 1-m long carbon block or on a 1m long tungsten collimator with jaws fully closed. An attenuation by a factor 100 is reached within 15 m behind a tungsten block, while the same takes 60 m for a carbon block. High Z materials are therefore used for absorption while much more robust, low Z materials are used for diluting any impacting beam.

Single collimators cannot achieve very high efficiencies. The limited absorption power of a single collimator is further reduced by out-scattering from the jaw surface. Particles diffuse slowly to large amplitudes in storage rings. They hit the collimator jaw as limiting aperture with small transverse offsets (impact parameter), in the order of a few 100 nm to a few  $\mu\text{m}$  [5]. Particles experience multiple Coulomb scattering in the jaw and many can escape through the collimator

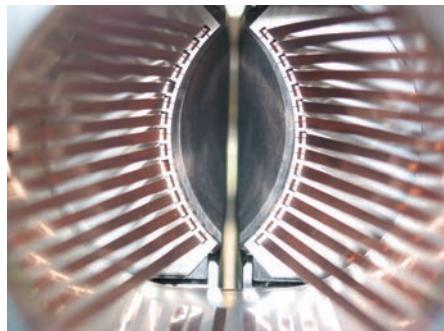


Figure 5: Photograph of a collimator gap defined by two parallel, flat jaws. The image currents of the beam are guided by rf fingers.

surface before the whole jaw length is traversed. Collimation is an edge process and becomes a multi-turn process in storage rings, with more than one impact on the primary collimator.

Collimator jaws usually are the closest materials to the beam in order to fulfill their function. Being so close, the electro-magnetic fields of the beam interact with the surface of the collimator jaw. The jaws must be tapered and rf contacts must be properly included into the collimator design to guide the image currents. The magnitude of the wakefields depend on the resistivity of the jaw material and the distance from the beam. For short bunches, dipolar deflections can be induced and were indeed observed and characterized for the SLAC linac [6]. For long bunches (like the LHC) the resistive wall impedance  $Z_{\perp}$  can be dominated by collimator-induced impedance ( $Z_{\perp} \propto 1/\text{gap}^3$  for a two-sided collimator) [7, 8]. For low frequencies the classical thick wall theory is modified, so-called “inductive bypass” (Secs.3.2.2,3.2.4).

The collimator volumes can in bad circumstances lead to trapped modes and collimator heating. Modern 3D simulations are applied to calculate any possible trapped modes and to address them by either modifying the collimator design or by placing ferrites for absorption. So-called rf fingers are required to guide the image currents of the beam and to prevent discontinuities. An example of a collimator gap and rf fingers is shown in Fig.5.

**Collimator-induced phase space cuts** Beam collimation is defined with normalized

coordinates  $z_n, z'_n$  ( $z = x$  or  $y$ ):

$$z_n = z / \sqrt{\beta_z \epsilon_z} \quad (1)$$

$$z'_n = \frac{\alpha_z \cdot z + \beta_z \cdot z'}{\sqrt{\beta_z \epsilon_z}} \quad (2)$$

Ideally a particle describes a circle with constant normalized amplitude  $a_z = \sqrt{z_n^2 + z'^2}$ . However, in practice  $a_z$  is stochastically increased by diffusion processes: space charge, scattering (IBS, beam-gas, beam-beam, ...), rf noise, ground motion and other effects. A beam halo (beam tails) forms around the beam core. A two-sided collimator shall be placed at a normalized distance  $\pm n_1$  from the beam center. Then its jaws are set at  $\pm n_1 \cdot \sqrt{\beta_{z,0} \epsilon_z}$  where  $\beta_{z,0}$  is the on-momentum beta function ( $\delta = 0$ ). The following combinations of betatron amplitudes ( $a_z$ ) and momentum offsets ( $\delta$ ) have particles impacting on the collimator edge [9]:

$$\pm n_1 = a_z \sqrt{\frac{\beta_z(\delta)}{\beta_{z,0}}} + \delta \frac{D_z(\delta)}{\sqrt{\beta_{z,0} \epsilon_z}} \quad (3)$$

A pure betatron cut  $a_z = n_1$  is achieved when the collimator is placed at a location with zero dispersion and chromatically well-corrected optical functions (no dependence on  $\delta$ ). Typically one uses  $4\sigma_z < n_1 < A_z$  with  $A_z$  being the normalized machine aperture that shall be protected against primary beam loss. Locations with high beta functions are advantageous for reducing wakefield and impedance effects, relaxing mechanical tolerances, increasing impact parameters and lowering beam density in case of erroneous hits (better robustness against error cases).

Maximizing the dispersion makes the collimator sensitive to off-momentum losses.  $n_1$  can be increased and the collimator intercepts predominantly off-momentum particles, especially when another collimator at zero dispersion cuts the betatron halo. An optimum is achieved with a maximum normalized dispersion at the primary momentum collimator [11]:

$$\frac{1}{D} \frac{dD}{ds} = -\frac{\alpha}{\beta} \quad (4)$$

The chromatic correction of the optics is essential in order to guarantee phase space cuts at the foreseen depths. This is illustrated in Fig.6. The energy dependence of the dispersion and the beta function is shown for the location of an LHC momentum collimator. The resulting change of the normalized betatron cut  $a_x$  is shown as well. The

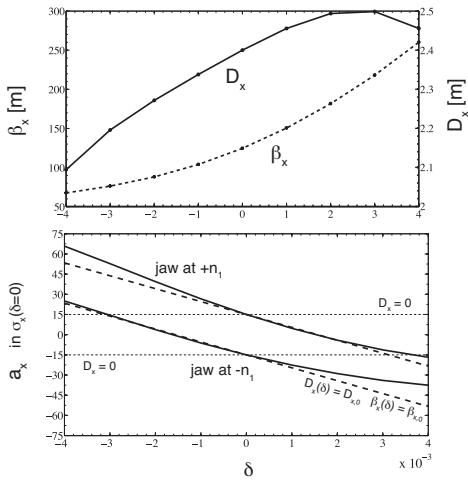


Figure 6: Energy dependence of beta and dispersion functions at a collimator location in the LHC (top) and its effect on the normalized betatron cut, Eq.(3) (bottom) [9].

large dispersion tilts the betatron cut versus energy offset  $\delta$ , while the chromatic change of the Twiss functions introduces a curvature in the betatron cut. Energy offsets can be large in the halo and chromatic effects can become limiting factors.

**System design guidelines** (i) The number of required primary collimators per plane depends on the details of the beam and the accelerator. A stored beam in a circular accelerator features a slow diffusion of primary beam halo that appears locally at the limiting aperture bottleneck. A single primary collimator (in the extreme case with one jaw) is sufficient to establish primary collimation of beam halo in a given plane at  $n_1$ . Distributed losses can also arise from synchrotron radiation, beam gas induced ionization of partially ionized beams, machine errors and other sources. A large number of distributed collimators or masks is required to intercept these losses. Finally, single pass accelerators and transfer lines require several primary collimators distributed in phase advance for constraining the maximum extent of beam halo [10]. The minimal system requires two two-sided primary collimators per plane, separated by  $90^\circ$  in phase advance. This would constrain the extent of beam halo to  $\sqrt{2} n_1$ . More primary collimators separated by less phase advance can reduce this maximum extent further.

### Sec.3.3: PARTICLE-MATTER INTERACTION

(ii) Gaps for a given  $n_1$  shall be maximized. This requires locations with a high beta function for betatron collimators and a high dispersion for momentum collimators. This guideline reduces wakefield and impedance effects, relaxes mechanical tolerances for collimators, increases impact parameters and lowers beam density in case of erroneous hits (better robustness against error cases).

(iii) A betatron collimation section shall be followed by a momentum collimation section, if high efficiency is important. The interactions in betatron collimators produce primary off-momentum particles.

(iv) Collimators can become highly radioactive. Large transverse space should be foreseen to provide possibilities for shielding and/or optimized handling.

**Multi-stage collimation systems** Highly efficient collimation systems rely on several stages. The classical two-stage collimation system [11, 12, 13] relies on a primary collimator for scattering and secondary collimators for absorption. The recent design at the LHC [14] has extended the two stages into four stages. Stages are defined by different normalized settings to the beam. For example, the primary collimators are closest to the beam and have a distance of  $n_1\sigma$  from the beam center. Defining the normalized setting of primary, secondary and tertiary collimators as  $n_1, n_2, n_3$  we have the following condition:

$$n_1 < n_2 < n_3 < A_z \quad (5)$$

where  $A_z$  is the normalized machine aperture as introduced above. The system designed for the LHC at 7 TeV is illustrated in Fig.7.

The values for  $n_1, n_2, n_3$  must be chosen such that the critical aperture  $A_z$  is effectively shadowed.  $n_1$  must be large enough not to reduce the beam lifetime or the transmission, normally  $n_1 > 4\sigma$ . The difference  $\Delta x_1 = n_2 - n_1$  is called secondary retraction,  $\Delta x_2 = n_3 - n_2$  is the tertiary retraction. It is important that the multi-stage hierarchy is fulfilled at all times. The retraction values must therefore be large enough to allow for drifts of the beam center and changes in beam size due to variations in the betatron function [1]. The impact of orbit and betatron errors are illustrated in Fig.8. The collimation system must be designed in coherence with a machine tolerance budget for proper functioning.

The one-dimensional betatron collimation system with two-stages has been fully described

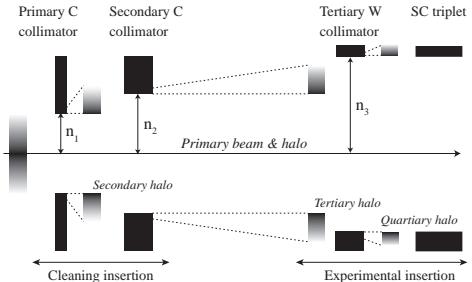


Figure 7: Illustration of multi-stage cleaning as used for the LHC. The primary collimators intercept the primary beam halo and leak a secondary halo. The secondary collimators intercept the secondary halo and leak a tertiary halo that is intercepted by tertiary collimators. A fourth stage also exists, but is not shown here.

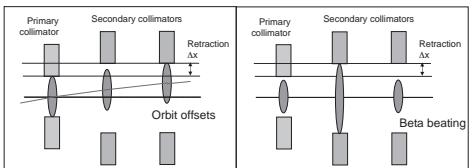


Figure 8: Illustration of machine errors (left: orbit error; right: beta beating) that can affect the collimation hierarchy in normalized phase space and reduce cleaning efficiency.

analytically [11, 12, 13]. An optimal efficiency can be achieved by placing two secondary collimators at unique phase advance locations behind the primary collimator. This is illustrated in Fig.9. In normalized phase space a particle describes a circle of amplitude  $n_1$ . At max  $z_n$  it is intercepted by a primary collimator which is set to  $n_1$  (position 1 in Fig.9). The particle receives a kick that increases its normalized amplitude to  $n_2$ . At the exit of the primary collimator it has position 2 in Fig.9. The particle must then be transported through some phase advance before it again reaches max  $z_n$  and can be intercepted by a secondary collimator set to  $n_2$  (position 3 in Fig.9). The phase advance depends on the sign of the kick received in the primary collimator.

The optimal phase advance  $\mu_{opt}$  from the exit of the primary collimator to the first primary collimator is calculated to be

$$\cos \mu_{opt} = \frac{n_1}{n_2} \quad (6)$$

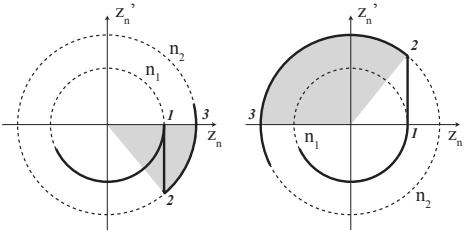


Figure 9: Illustration of collimator induced particle kicks and optimal phase advance to a secondary collimator. See discussion in text.

The two secondary collimators should then be placed at phase advances  $\mu_1$  and  $\mu_2$  downstream of the primary collimator:

$$\mu_1 = \mu_{opt} \quad \text{and} \quad \mu_2 = \pi - \mu_{opt} \quad (7)$$

For a realistic system design, however, it is not sufficient to consider a 1D collimation system as sketched above [15]. A particle impacting on a horizontal collimator will be scattered both in horizontal and vertical directions. It is therefore required to design a more sophisticated system. The maximum extent  $A_1^{\max}$  of the primary halo is constrained by the primary collimators. For two orthogonal primary collimators at  $n_1$  (usually in horizontal and vertical orientation) [13]:

$$A_1^{\max} = \sqrt{2} n_1 \approx 1.41 n_1 \quad (8)$$

For three primary collimators with  $45^\circ$  coverage (horizontal, vertical, skew)  $A_1^{\max}$  becomes:

$$A_1^{\max} = n_1 / \cos(\pi/8) \approx 1.08 n_1 \quad (9)$$

It is therefore advantageous to design a system with three primary collimators. The need for the skew collimator must be assessed with a study on the expected beam halo densities: skew collimators are effective in cases where there are many halo particles with simultaneously large horizontal and vertical amplitudes.

The optimal number, orientation and phase advance locations of secondary collimators can be calculated with full collimation theory. The extension of secondary beam halo is quoted in Tab.1 as a function of the number  $N_{\text{sec}}$  of secondary collimators per primary collimator.

The secondary beam halo can be efficiently constrained by placing four secondary collimators per primary collimator. The betatron collimation system then consists of three primary and 12 secondary collimators (in total 15 collimators with

Table 1: The extension  $A_2^{\max}$  of secondary beam halo is listed for three primary collimators and for different numbers  $N_{\text{sec}}$  of secondary collimators per primary. Collimators with two flat, parallel jaws are assumed. The last column shows a realistic example. [13]

| $N_{\text{sec}}$ | $A_2^{\max}$   | $A_2^{\max}$<br>( $n_1 = 6, n_2 = 7$ ) |
|------------------|--|--|
| 3                | $4n_2^2 - 3n_1^2$                                      | 9.4                                    |
| 4                | $2n^2 - n_1^2$   | 7.9                                    |
| 8                | $\frac{4n_2^2 - 2n_1^2 + \sqrt{2n_1^2}}{2 + \sqrt{2}}$ | 7.2                                    |
| $\infty$         | $n_2$  | 7.0                                    |

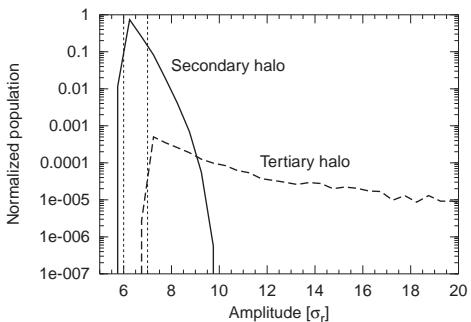


Figure 10: Extension of secondary and tertiary beam halo at a zero dispersion location, as simulated for the LHC multi-stage collimation system at 7 TeV with  $n_1 = 6$  and  $n_2 = 7$ .

30 jaws). The system is only efficient if the secondary collimators are placed at the correct phase advances from the primary collimator. The best solution requires a modulated optics with unequal horizontal and vertical phase advance. Detailed solutions and tables for optimal theoretical phase advance locations are listed in [13]. The simulated shapes of secondary and tertiary halos are shown in Fig.10 for 7 TeV LHC collimation [14], here at a location of zero dispersion.

It is noted that the quoted extension of the secondary beam halo is correct for particles that are scattered in a jaw, but do not lose significant amounts of energy. This is not fully realistic. Even at multi-TeV beam energies some particles will lose a significant energy in the primary jaw without being stopped. These particles will escape as off-momentum particles. At locations of high dispersion the particles' amplitude will extend beyond  $A_2^{\max}$ . This is shown in Fig.11

### Sec.3.3: PARTICLE-MATTER INTERACTION

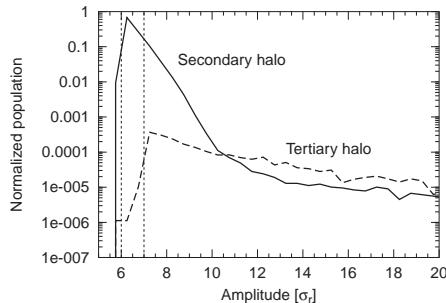


Figure 11: The same simulation as shown in Fig.10, but now at a location of maximum dispersion in the LHC. It is seen that the secondary halo contains off-energy protons. These originate mostly from single-diffractive scattering in the primary collimator jaw.

for the same case as in Fig.10. It is therefore important to append betatron collimators to an off-momentum collimation section.

#### Simulation of beam halo and collimation

Modern collimation systems are designed based on simulations of optics, phase space coverage, impedance, beam halo and cleaning efficiency:

(i) Accelerator optics programs are used to design the collimation insertion and its optics, minimizing the extent of downstream halo amplitudes [16].

(ii) Special routines are used to constrain impedance from collimators at the same time, trying to place collimators at the locations of larger beta values.

(iii) Tracking programs are used to track up to 20 million halo particles for several 100 turns, while simulating the particle interaction in collimator materials [17, 18]. The calculated efficiency is passed back into (i) for a next design iteration. Simulations with reduced numbers of collimators identify the most and least efficient collimators, allowing further optimization.

(iv) Special aperture analysis programs pass the halo found in (iii) through a model of the machine hardware to predict the loss patterns. Particular loss peaks in sensitive equipment are identified and local collimators can be placed for further enhancing collimation efficiency.

An example of multi-turn tracking for vertical halo in the LHC is shown in Fig.12. These simulation data yielded the results of Figs.10 and 11.

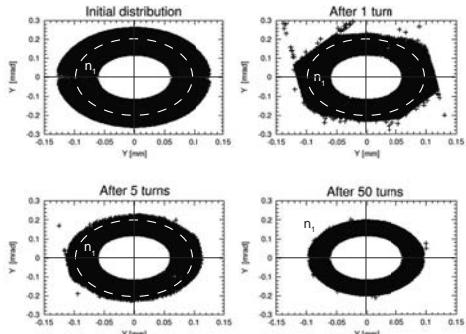


Figure 12: Simulation of multi-turn cleaning in the LHC storage ring. The vertical phase space is shown at the start, after 1, 5 and 50 turns. Primary collimation at  $n_1$  is indicated. The collimator jaw surfaces show up as straight lines which define the scattering locations (see 1 turn).

**Performance** The inefficiency  $\eta_c$  of a collimation system is defined as the ratio between the number  $N_{\text{leak}}$  of particles that leak out with  $a_z > a_z^{\text{cut}}$  and the number  $N_{\text{impact}}$  of impacting particles:

$$\eta_c = \frac{N_{\text{leak}}(a_z > a_z^{\text{cut}})}{N_{\text{impact}}} \quad (10)$$

with  $a_z^{\text{cut}} > n_1$ . The value of  $a_z^{\text{cut}}$  is given by the available machine aperture and is often around 10 sigma. Modern collimation systems can reach quite low inefficiencies with  $\eta_c$  in the range of  $10^{-2}$  (1%) to  $10^{-4}$  (0.01%). Efficiency is defined as  $\eta = 1 - \eta_c$  and is in the range of 99% to 99.99%.

Inefficiency is not sufficient to characterize the performance of collimation. Large amplitude particles are not lost at one location but are spread over some dilution length  $L_{\text{dil}}$ . A local cleaning inefficiency  $\tilde{\eta}_c$  [1] is defined as

$$\tilde{\eta}_c = \frac{\eta_c}{L_{\text{dil}}} \quad (11)$$

with units of 1/m. Dilution is not uniform and simulations are used to predict the local cleaning inefficiency  $\tilde{\eta}_c$  along the accelerator. Local cleaning inefficiency is minimized by reducing global inefficiency (overall leakage) and maximizing dilution. It is crucial to work on both aspects for achieving best performance.

A maximum local cleaning inefficiency  $\max[\tilde{\eta}_c]$  is defined over all critical locations. For example, in a super-conducting storage ring

$\max[\tilde{\eta}_c]$  describes the peak loss per m in superconducting magnets. Alternatively, in a linac  $\max[\tilde{\eta}_c]$  may describe the peak loss per m in the regions that must be protected for hands-on maintenance.

The maximum allowed loss rate  $R_{\text{loss}}$  at the primary collimators then depends on the allowable maximum beam loss rate  $R_{\text{lim}}$  for critical locations and the maximum local inefficiency of the system [1]:

$$R_{\text{loss}} = \frac{R_{\text{lim}}}{\max[\tilde{\eta}_c]} \quad (12)$$

Considering a stored beam and assuming that particles are lost at collimators we can relate  $R_{\text{loss}} = \Delta N / \Delta T$  to the number of particles  $N_{\text{max}}$  and beam lifetime  $\tau_{\text{min}}$ :

$$\tau_{\text{min}} = -\frac{\Delta T}{\ln \left( 1 - \frac{R_{\text{loss}} \cdot \Delta T}{N_{\text{max}}} \right)} \approx \frac{N_{\text{max}}}{R_{\text{loss}}} \quad (13)$$

The maximum achievable beam intensity can be expressed as a function of the maximum local cleaning inefficiency, the minimum beam lifetime that must be sustained and the limit of beam loss in critical regions [1]:

$$N_{\text{max}} = \frac{\tau_{\text{min}} \cdot R_{\text{lim}}}{\max[\tilde{\eta}_c]} \quad (14)$$

This equation is used during the design phase to determine the required collimation performance once beam intensity, minimum beam lifetime and loss limits have been fixed. Similar equations can be introduced for single-pass accelerators.

**Beam-based setup of collimation** Collimators must be centered around the beam with an accuracy that is a fraction of the collimator retraction. Tolerances for collimator settings can then be in the range of a few 10's of  $\mu\text{m}$ . However, the exact beam position and size are not known a priori with this accuracy. Collimators are therefore set up in a beam-based process. This beam-based procedure differs for one-pass or stored beams.

In a single pass accelerator or in transfer lines a collimator jaw is moved through the beam while observing the transmission of beam and downstream showers due to beam impact on the collimator material. In a storage rings a reference collimator (most often a primary collimator) is used to define a betatron cut in normalized phase space. Assuming zero dispersion, the same phase space

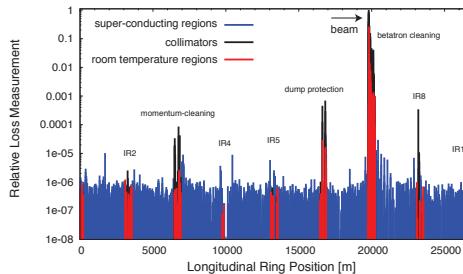


Figure 13: Example for a measurement of collimation performance in the LHC at 450 GeV. The data shows peak integrated losses over 1.3 s. A beam loss is provoked for beam 1 in the horizontal plane (emittance blowup). Losses are normalized to the peak loss in the ring. The peak loss appears as expected at the betatron collimators and falls off exponentially over the betatron cleaning insertion. Leakage around the ring is measured. The measurement resolution is limited by noise in the beam loss monitors (6 orders of magnitudes below the peak loss).

cut is present all around the ring after phase space mixing (particles oscillating around the closed orbit, sweeping around the whole allowed phase space volume). The second jaw of the reference collimator can then be moved to the same cut. A sudden spike in beam loss measured downstream of the collimator is used to detect the halo edge. Successively all collimators around the ring are set up to the same cut in normalized phase space. In the end all jaws are centered on the beam and any beta variations have been calibrated.

**Measurement of collimation performance** Collimation performance can be measured if a distributed beam loss measurement (BLM) system has been installed around the ring. Ideally beam loss is measured at all collimators, all quadrupoles (here the beta functions are maximal) and other critical locations. An example measurement of collimation performance in the LHC at 450 GeV is shown in Fig.13. The procedure for such a measurement is described:

(i) The measurement shall not disturb the orbit nor the beta functions, as this would decrease the cleaning efficiency. Therefore one induces a strong diffusion process that rapidly increases the beam emittance. In a storage ring one can move the beam onto a resonance, for example the 1/3 resonance, or one can blow up individual bunches by gated noise excitation.

(ii) The integrated beam losses are monitored around the ring as the beam emittance is blown up.

(iii) The loss data is normalized to the highest loss all around the ring, which by construction should occur at a collimator.

Such a measurement is shown in Fig.13 [19] and is closely related to the local cleaning inefficiency defined above. Differences arise due to BLM response characteristics. The measured maximum ‘local cleaning inefficiency’ in a critical region (super-conducting magnets) is about  $2 \times 10^{-5}$  in Fig.13. It is seen that leakage is very small.

## References

- [1] R. Assmann, CERN-AB-2003-008 ADM (2003)
- [2] N. Mokhov et al, FERMILAB-Conf-03/220
- [3] M. Seidel, DESY 94-103 (1994)
- [4] P.J. Bryant et al, CERN SL/93-15(AP) (1993)
- [5] R. Assmann et al, CERN LHC PR 592 (2002)
- [6] D. Onoprienko et al, SLAC-PUB-10192 (2002)
- [7] E. Metral et al, CERN LHC PR 1015 (2007)
- [8] F. Zimmermann, CERN-AB-Note-2006-007
- [9] C. Bracco, R. Assmann, CERN-ATS-2009-033
- [10] S. Di Mitri, PRST-AB 13, 052801 (2010)
- [11] P.J. Bryant, E. Klein, CERN SL/92-40(AP) (1992)
- [12] T. Trenkler, J.B. Jeanneret, PA 50 (1995) 287
- [13] J.B. Jeanneret, PRST-AB 1, 081001 (1998)
- [14] R. Assmann et al, CERN LHC PR 919 (2006)
- [15] P.J. Bryant, CERN SL/92-24(AP) (1992)
- [16] D. Kaltchev et al, CERN LHC PR 134 (1997)
- [17] G. Robert-Demolaize et al, CERN-AB-2005-033 (2005)
- [18] S. Redaelli et al, CERN LHC PR 938 (2006)
- [19] D. Wollmann et al, IPAC 10

## 3.3.12 Atomic and Nuclear Properties of Materials [1]

The table on the next page is abridged from pdg.lbl.gov/AtomicNuclearProperties by D.E. Groom (2007). See web pages for more detail about entries in this table including chemical formulae, and for several hundred other entries. Quantities in parentheses are for NTP ( $20^\circ\text{C}$  and 1 atm), and square brackets indicate quantities evaluated at STP. Boiling points are at 1 atm. Refractive indices  $n$  are evaluated at the sodium D line blend (589.2 nm); values  $\gg 1$  in brackets are for  $(n - 1) \times 10^6$  (gases).

## References

- [1] K. Nakamura et al, (Particle Data Group), J. Phys. G37, 075021 (2010), and 2011 partial update for the 2012 edition

| Material  | Z  | A                | d(Z/A)  | Nucl.coll.<br>length $\lambda_T$<br>{g cm <sup>-2</sup> } | Nucl.inter.<br>length $\lambda_f$<br>{g cm <sup>-2</sup> } | Rad.len.<br>$X_0$<br>{g cm <sup>-2</sup> } | $dE/dx _{\min}$<br>MeV<br>{g <sup>-1</sup> cm <sup>2</sup> } | Density<br>{g cm <sup>-3</sup> }<br>(g l <sup>-1</sup> ) | Melting<br>point<br>(K) | Boiling<br>point<br>(K) | Refract.<br>index<br>(@ Na D) |
|---|----|------------------|---------|---|--|--|--|--|-------------------------|-------------------------|-------------------------------|
| H <sub>2</sub>  | 1  | 1.00794(7)       | 0.99212 | 42.8  | 52.0   | 63.04                                      | (4.103)  | 0.071(0.084)   | 13.81                   | 20.28                   | 1.11[132.]                    |
| D <sub>2</sub>  | 1  | 2.01410177803(8) | 0.49650 | 51.3  | 71.8   | 125.97                                     | (2.053)  | 0.169(0.168)   | 18.7                    | 23.65                   | 1.11[138.]                    |
| He  | 2  | 4.002602(2)      | 0.49967 | 51.8  | 71.0   | 94.32                                      | (1.937)  | 0.125(0.166)   |                         | 4.220                   | 1.02[35.0]                    |
| Li  | 3  | 6.941(2)         | 0.43221 | 52.2  | 71.3   | 82.78                                      | 1.639  | 0.534  | 453.6                   | 1615.                   |                               |
| Be  | 4  | 9.012182(3)      | 0.44384 | 55.3  | 77.8   | 65.19                                      | 1.595  | 1.848  | 1560.                   | 2744.                   |                               |
| C diamond   | 6  | 12.0107(8)       | 0.49955 | 59.2  | 85.8   | 42.70                                      | 1.725  | 3.520  |                         |                         | 2.42                          |
| C graphite  | 6  | 12.0107(8)       | 0.49955 | 59.2  | 85.8   | 42.70                                      | 1.742  | 2.210  |                         |                         |                               |
| N <sub>2</sub>  | 7  | 14.0067(2)       | 0.49976 | 61.1  | 89.7   | 37.99                                      | (1.825)  | 0.807(1.165)   | 63.15                   | 77.29                   | 1.20[298.]                    |
| O <sub>2</sub>  | 8  | 15.9994(3)       | 0.50002 | 61.3  | 90.2   | 34.24                                      | (1.801)  | 1.141(1.332)   | 54.36                   | 90.20                   | 1.22[271.]                    |
| F <sub>2</sub>  | 9  | 18.9984032(5)    | 0.47372 | 65.0  | 97.4   | 32.93                                      | (1.676)  | 1.507(1.580)   | 53.53                   | 85.03                   | [195.]                        |
| Ne  | 10 | 20.1797(6)       | 0.49555 | 65.7  | 99.0   | 28.93                                      | (1.724)  | 1.204(0.839)   | 24.56                   | 27.07                   | 1.09[67.1]                    |
| Al  | 13 | 26.9815386(8)    | 0.48181 | 69.7  | 107.2  | 24.01                                      | 1.615  | 2.699  | 933.5                   | 2792.                   |                               |
| Si  | 14 | 28.0855(3)       | 0.49848 | 70.2  | 108.4  | 21.82                                      | 1.664  | 2.329  | 1687.                   | 3538.                   | 3.95                          |
| Cl <sub>2</sub>   | 17 | 35.453(2)        | 0.47951 | 73.8  | 115.7  | 19.28                                      | (1.630)  | 1.574(2.980)   | 171.6                   | 239.1                   | [773.]                        |
| Ar  | 18 | 39.948(1)        | 0.45059 | 75.7  | 119.7  | 19.55                                      | (1.519)  | 1.396(1.662)   | 83.81                   | 87.26                   | 1.23[281.]                    |
| Ti  | 22 | 47.867(1)        | 0.45961 | 78.8  | 126.2  | 16.16                                      | 1.477  | 4.540  | 1941.                   | 3560.                   |                               |
| Fe  | 26 | 55.845(2)        | 0.46557 | 81.7  | 132.1  | 13.84                                      | 1.451  | 7.874  | 1811.                   | 3134.                   |                               |
| Cu  | 29 | 63.546(3)        | 0.45636 | 84.2  | 137.3  | 12.86                                      | 1.403  | 8.960  | 1358.                   | 2835.                   |                               |
| Ge  | 32 | 72.64(1)         | 0.44053 | 86.9  | 143.0  | 12.25                                      | 1.370  | 5.323  | 1211.                   | 3106.                   |                               |
| Sn  | 50 | 118.710(7)       | 0.42119 | 98.2  | 166.7  | 8.82                                       | 1.263  | 7.310  | 505.1                   | 2875.                   |                               |
| Xe  | 54 | 131.293(6)       | 0.41129 | 100.8   | 172.1  | 8.48                                       | (1.255)  | 2.953(5.483)   | 161.4                   | 165.1                   | 1.39[701.]                    |
| W   | 74 | 183.84(1)        | 0.40252 | 110.4   | 191.9  | 6.76                                       | 1.145  | 19.300   | 3695.                   | 5828.                   |                               |
| Pt  | 78 | 195.084(9)       | 0.39983 | 112.2   | 195.7  | 6.54                                       | 1.128  | 21.450   | 2042.                   | 4098.                   |                               |
| Au  | 79 | 196.966569(4)    | 0.40108 | 112.5   | 196.3  | 6.46                                       | 1.134  | 19.320   | 1337.                   | 3129.                   |                               |
| Pb  | 82 | 207.2(1)         | 0.39575 | 114.1   | 199.6  | 6.37                                       | 1.122  | 11.350   | 600.6                   | 2022.                   |                               |
| U   | 92 | [238.02891(3)]   | 0.38651 | 118.6   | 209.0  | 6.00                                       | 1.081  | 18.950   | 1408.                   | 4404.                   |                               |
| Air (dry, 1 atm)  |    | 0.49919          | 61.3    | 90.1  | 36.62  | (1.815)                                    | (1.205)  |  | 78.80                   |                         |                               |
| Shielding concrete  |    | 0.50274          | 65.1    | 97.5  | 26.57  | 1.711                                      | 2.300  |  |                         |                         |                               |
| Borosilicate glass (Pyrex)  |    | 0.49707          | 64.6    | 96.5  | 28.17  | 1.696                                      | 2.230  |  |                         |                         |                               |
| Lead glass  |    | 0.42101          | 95.9    | 158.0   | 7.87   | 1.255                                      | 6.220  |  |                         |                         |                               |
| Standard rock   |    | 0.50000          | 66.8    | 101.3   | 26.54  | 1.688                                      | 2.650  |  |                         |                         |                               |
| Methane (CH <sub>4</sub> )  |    | 0.62334          | 54.0    | 73.8  | 46.47  | (2.417)                                    | (0.667)  | 90.68  | 111.7                   | [444.]                  |                               |
| Ethane (C <sub>2</sub> H <sub>6</sub> )   |    | 0.59861          | 55.0    | 75.9  | 45.66  | (2.304)                                    | (1.263)  | 90.36  | 184.5                   |                         |                               |
| Propane (C <sub>3</sub> H <sub>8</sub> )  |    | 0.58962          | 55.3    | 76.7  | 45.37  | (2.262)                                    | 0.493(1.868)   | 85.52  | 231.0                   |                         |                               |
| Butane (C <sub>4</sub> H <sub>10</sub> )  |    | 0.59497          | 55.5    | 77.1  | 45.23  | (2.278)                                    | (2.489)  | 134.9  | 272.6                   |                         |                               |
| Octane (C <sub>8</sub> H <sub>18</sub> )  |    | 0.57778          | 55.8    | 77.8  | 45.00  | 2.123                                      | 0.703  | 214.4  | 398.8                   |                         |                               |
| Paraffin (CH <sub>3</sub> (CH <sub>2</sub> ) <sub>n</sub> ≈ <sub>23</sub> CH <sub>3</sub> ) |    | 0.57275          | 56.0    | 78.3  | 44.85  | 2.088                                      | 0.930  |  |                         |                         |                               |
| Nylon (type 6, 6/6)   |    | 0.54790          | 57.5    | 81.6  | 41.92  | 1.973                                      | 1.18   |  |                         |                         |                               |
| Polycarbonate (Lexan)   |    | 0.52697          | 58.3    | 83.6  | 41.50  | 1.886                                      | 1.20   |  |                         |                         |                               |
| Polyethylene ([-CH <sub>2</sub> CH <sub>2</sub> ] <sub>n</sub> )                            |    | 0.57034          | 56.1    | 78.5  | 44.77  | 2.079                                      | 0.89   |  |                         |                         |                               |
| Polyethylene terephthalate (Mylar)  |    | 0.52037          | 58.9    | 84.9  | 39.95  | 1.848                                      | 1.40   |  |                         |                         |                               |
| Polyimide film (Kapton)   |    | 0.51264          | 59.2    | 85.5  | 40.58  | 1.820                                      | 1.42   |  |                         |                         |                               |
| Polymethylmethacrylate (acrylic)  |    | 0.53937          | 58.1    | 82.8  | 40.55  | 1.929                                      | 1.19   |  |                         |                         | 1.49                          |
| Polypropylene   |    | 0.55998          | 56.1    | 78.5  | 44.77  | 2.041                                      | 0.90   |  |                         |                         |                               |
| Polystyrene ([C <sub>6</sub> H <sub>5</sub> CHCH <sub>2</sub> ] <sub>n</sub> )              |    | 0.53768          | 57.5    | 81.7  | 43.79  | 1.936                                      | 1.06   |  |                         |                         | 1.59                          |
| Polytetrafluoroethylene (Teflon)  |    | 0.47992          | 63.5    | 94.4  | 34.84  | 1.671                                      | 2.20   |  |                         |                         |                               |
| Polyvinyltoluene  |    | 0.54141          | 57.3    | 81.3  | 43.90  | 1.956                                      | 1.03   |  |                         |                         | 1.58                          |
| Aluminum oxide (sapphire)   |    | 0.49038          | 65.5    | 98.4  | 27.94  | 1.647                                      | 3.970  | 2327.  | 3273.                   | 1.77                    |                               |
| Barium fluoride (BaF <sub>2</sub> )   |    | 0.42207          | 90.8    | 149.0   | 9.91   | 1.303                                      | 4.893  | 1641.  | 2533.                   | 1.47                    |                               |
| Bismuth germanate (BGO)   |    | 0.42065          | 96.2    | 159.1   | 7.97   | 1.251                                      | 7.130  | 1317.  |                         |                         | 2.15                          |
| Carbon dioxide gas (CO <sub>2</sub> )   |    | 0.49989          | 60.7    | 88.9  | 36.20  | 1.819                                      | (1.842)  |  |                         |                         | [449.]                        |
| Solid carbon dioxide (dry ice)  |    | 0.49989          | 60.7    | 88.9  | 36.20  | 1.787                                      | 1.563  | Sublimes at 194.7 K                                      |                         |                         |                               |
| Cesium iodide (CsI)   |    | 0.41569          | 100.6   | 171.5   | 8.39   | 1.243                                      | 4.510  | 894.2  | 1553.                   | 1.79                    |                               |
| Lithium fluoride (LiF)  |    | 0.46262          | 61.0    | 88.7  | 39.26  | 1.614                                      | 2.635  | 1121.  | 1946.                   | 1.39                    |                               |
| Lithium hydride (LiH)   |    | 0.50321          | 50.8    | 68.1  | 79.62  | 1.897                                      | 0.820  | 965.   |                         |                         |                               |
| Lead tungstate (PbWO <sub>4</sub> )   |    | 0.41315          | 100.6   | 168.3   | 7.39   | 1.229                                      | 8.300  | 1403.  |                         |                         | 2.20                          |
| Silicon dioxide (SiO <sub>2</sub> , fused quartz)   |    | 0.49930          | 65.2    | 97.8  | 27.05  | 1.699                                      | 2.200  | 1986.  | 3223.                   | 1.46                    |                               |
| Sodium chloride (NaCl)  |    | 0.55509          | 71.2    | 110.1   | 21.91  | 1.847                                      | 2.170  | 1075.  | 1738.                   | 1.54                    |                               |
| Sodium iodide (NaI)   |    | 0.42697          | 93.1    | 154.6   | 9.49   | 1.305                                      | 3.667  | 933.2  | 1577.                   | 1.77                    |                               |
| Water (H <sub>2</sub> O)  |    | 0.55509          | 58.5    | 83.3  | 36.08  | 1.992                                      | 1.000(0.756)   | 273.1  | 373.1                   | 1.33                    |                               |
| Silica aerogel  |    | 0.50093          | 65.0    | 97.3  | 27.25  | 1.740                                      | 0.200  | (0.03 H <sub>2</sub> O, 0.97 SiO <sub>2</sub> )          |                         |                         |                               |

### Sec.3.3: PARTICLE-MATTER INTERACTION

| Material       | Dielectric constant ( $\kappa = \epsilon/\epsilon_0$ )<br>( $\kappa$ is $(\kappa-1) \times 10^6$<br>for gas) | Young's modulus<br>[ $10^6$ psi] | Coeff. of thermal expansion<br>[ $10^{-6}$ cm/cm-°C] | Specific heat<br>[cal/g-°C] | Electrical resistivity<br>[ $\mu\Omega\text{cm}(@^\circ\text{C})$ ] | Thermal conductivity<br>[cal/cm-°C-sec] |
|----------------|--|----------------------------------|--|-----------------------------|---|---|
| H <sub>2</sub> | (253.9)  | —                                | —  | —                           | —   | —                                       |
| He             | (64)   | —                                | —  | —                           | —   | —                                       |
| Li             | —  | —                                | 56   | 0.86                        | 8.55(0°)  | 0.17                                    |
| Be             | —  | 37                               | 12.4   | 0.436                       | 5.885(0°)   | 0.38                                    |
| C              | —  | 0.7                              | 0.6-4.3  | 0.165                       | 1375(0°)  | 0.057                                   |
| N <sub>2</sub> | (548.5)  | —                                | —  | —                           | —   | —                                       |
| O <sub>2</sub> | (495)  | —                                | —  | —                           | —   | —                                       |
| Ne             | (127)  | —                                | —  | —                           | —   | —                                       |
| Al             | —  | 10                               | 23.9   | 0.215                       | 2.65(20°)   | 0.53                                    |
| Si             | 11.9   | 16                               | 2.8-7.3  | 0.162                       | —   | 0.20                                    |
| Ar             | (517)  | —                                | —  | —                           | —   | —                                       |
| Ti             | —  | 16.8                             | 8.5  | 0.126                       | 50(0°)  | —                                       |
| Fe             | —  | 28.5                             | 11.7   | 0.11                        | 9.71(20°)   | 0.18                                    |
| Cu             | —  | 16                               | 16.5   | 0.092                       | 1.67(20°)   | 0.94                                    |
| Ge             | 16.0   | —                                | 5.75   | 0.073                       | —   | 0.14                                    |
| Sn             | —  | 6                                | 20   | 0.052                       | 11.5(20°)   | 0.16                                    |
| Xe             | —  | —                                | —  | —                           | —   | —                                       |
| W              | —  | 50                               | 4.4  | 0.032                       | 5.5(20°)  | 0.48                                    |
| Pt             | —  | 21                               | 8.9  | 0.032                       | 9.83(0°)  | 0.17                                    |
| Pb             | —  | 2.6                              | 29.3   | 0.038                       | 20.65(20°)  | 0.083                                   |
| U              | —  | —                                | 36.1   | 0.028                       | 29(20°)   | 0.064                                   |

# Chapter 4. OPERATIONAL CONSIDERATIONS

## 4.1 LUMINOSITY

*M.A. Furman, M.S. Zisman, LBNL[1]*

**Definition** When two bunches (+ and -) having  $N_{\pm}$  particles and distributions  $\rho_{\pm}(\mathbf{x}, t)$  (normalized by  $\int d^3 \mathbf{x} \rho_{\pm}(\mathbf{x}, t) = N_{\pm}$ ) collide, the single-collision luminosity  $\mathcal{L}_{sc}$  for a particular reaction is defined as the number of events produced per unit cross section  $\sigma$  for this reaction, and is given by the overlap integral [2, 3]

$$\mathcal{L}_{sc} = \frac{1}{c} \int d^3 \mathbf{x} dt \rho_+(\mathbf{x}, t) \rho_-(\mathbf{x}, t) \times \sqrt{c^2(\mathbf{v}_+ - \mathbf{v}_-)^2 - (\mathbf{v}_+ \times \mathbf{v}_-)^2} \quad (1)$$

where  $\mathbf{v}_+(\mathbf{v}_-)$  is the common velocity of all the particles in bunch +(-). Eq.(1) is a relativistic invariant, has dimensions of 1/area, and is valid for arbitrary velocities  $\mathbf{v}_{\pm}$  (see [4] for generalization to the case when velocity distributions are not homogeneous).

For a storage ring collider in which the bunches are spaced by a distance  $s_B$ , bunches collide periodically with frequency  $f_c = \beta c / s_B$  yielding  $\dot{N}$  reaction events per unit time. For a linear collider,  $f_c = (\text{repetition rate}) \times (\text{number of bunches per bunch train})$ . The peak luminosity is given by  $\mathcal{L} = \dot{N} / \sigma = f_c \mathcal{L}_{sc}$  [5]. It is traditionally expressed in cgs units,  $\text{cm}^{-2} \text{s}^{-1}$ .

Tab.1 gives expressions for  $\mathcal{L}$  in various situations for head-on  $e^+e^-$  collisions<sup>1</sup> and  $\sigma_z$  small compared to  $\beta_x^*, \beta_y^*$ . These expressions are valid even with nonzero dispersion at the IP, unless otherwise noted. For initial estimates using Tab.1, we assume the nominal emittances and beam sizes, but these values generally evolve in time owing to the beam-beam force, and the luminosity should be modified accordingly [6].

<sup>1</sup>For particle species other than  $e^+e^-$  these expressions remain valid if one replaces the classical electron radius  $r_e$  by the appropriate value, provided the colliding particle species have equal mass.

In the  $y$  plane, the beam-beam tune shift parameter of an on-axis particle in the positron beam due to its interaction with the opposing electron beam is

$$\xi_{y,+} = \frac{r_e N_- \beta_{y,+}^*}{2\pi\gamma_+ \sigma_{y,-}^* (\sigma_{x,-}^* + \sigma_{y,-}^*)} \quad (2)$$

Expressions for the remaining 3 tune shift parameters are obtained by  $x \leftrightarrow y$  and/or  $+ \leftrightarrow -$ .

**Transparency symmetry** In a two-ring  $e^+e^-$  collider, beam parameters need not be identical in both rings. In the initial design of the two asymmetric B factories PEP-II [6] and KEKB [7], it was suggested to restrict the available parameter space [8, 9, 10] by choosing values to mimic the situation in a symmetric (single-ring) collider. The “transparency” conditions thus reached for two-ring colliders include: (i) pairwise equality of beam-beam tune shift parameters ( $\xi_{x,+} = \xi_{x,-}$ ;  $\xi_{y,+} = \xi_{y,-}$ ); (ii) pairwise equality of beam sizes ( $\sigma_{x,+}^* = \sigma_{x,-}^*$ ;  $\sigma_{y,+}^* = \sigma_{y,-}^*$ ); (iii) equality of tune modulation amplitudes associated with synchrotron oscillations ( $(\sigma_z \nu_s / \beta_{x,y}^*)_+ = (\sigma_z \nu_s / \beta_{x,y}^*)_-$ ; and sometimes (iv) equality of radiation damping decrements for the two rings. In particular, choosing parameters such that all four beam-beam parameters are equal is called “optimal coupling.” This case requires [6, 8]

$$\begin{aligned} \left( \frac{\sigma_y^*}{\sigma_x^*} \right)_+ &= \left( \frac{\sigma_y^*}{\sigma_x^*} \right)_- = \left( \frac{\beta_y^*}{\beta_x^*} \right)_+ = \left( \frac{\beta_y^*}{\beta_x^*} \right)_- \\ &= \left( \frac{\epsilon_y}{\epsilon_x} \right)_+ = \left( \frac{\epsilon_y}{\epsilon_x} \right)_- \equiv r \end{aligned} \quad (3)$$

Operational experience at PEP-II and KEKB [11], as well as at the Tevatron and HERA, has shown that transparency symmetry is not an essential requirement for stable operation; the observed beam-beam dynamics generally varies smoothly when conditions depart from this symmetry.

Table 1: Head-on luminosity expressions for short  $e^+e^-$  upright Gaussian bunches.

| Expression for $\mathcal{L}$   | Conditions for validity   |
|--|---|
| $\frac{N_+ N_- f_c}{2\pi \sqrt{(\sigma_{x,+}^{*2} + \sigma_{x,-}^{*2})(\sigma_{y,+}^{*2} + \sigma_{y,-}^{*2})}}$ | general   |
| $\frac{N_+ N_- f_c}{4\pi \sigma_x^* \sigma_y^*}$   | $\sigma_{x,+}^* = \sigma_{x,-}^* \equiv \sigma_x^*$ , $\sigma_{y,+}^* = \sigma_{y,-}^* \equiv \sigma_y^*$   |
| $K(1+r)\xi_y \left( \frac{EI}{\beta_y^*} \right)_{+,-}$  | $\sigma_{x,+}^* = \sigma_{x,-}^* \equiv \sigma_x^*$ , $\sigma_{y,+}^* = \sigma_{y,-}^* \equiv \sigma_y^*$ ,<br>$\xi_{x,+} = \xi_{x,-} \equiv \xi_x$ , $\xi_{y,+} = \xi_{y,-} \equiv \xi_y$  |
| $K(1+r) \left( \xi \frac{EI}{\beta_y^*} \right)_{+,-}$   | $\sigma_{x,+}^* = \sigma_{x,-}^* \equiv \sigma_x^*$ , $\sigma_{y,+}^* = \sigma_{y,-}^* \equiv \sigma_y^*$ ,<br>$\xi_{x,+} = \xi_{y,+} \equiv \xi_+$ , $\xi_{x,-} = \xi_{y,-} \equiv \xi_-$  |
| $\frac{N f_c \gamma \xi}{r_e \beta^*}$   | $\sigma_{x,+}^* = \sigma_{x,-}^* = \sigma_{y,+}^* = \sigma_{y,-}^*$ ,<br>$\beta_{x,+}^* = \beta_{x,-}^* = \beta_{y,+}^* = \beta_{y,-}^* \equiv \beta^*$ ,<br>$N_+ = N_- \equiv N$ , $E_+ = E_- \equiv E$  |
| $\frac{N^2 f_c}{4\pi \epsilon \beta^*}$  | $\epsilon_{x,+} = \epsilon_{x,-} = \epsilon_{y,+} = \epsilon_{y,-} \equiv \epsilon$ ,<br>$\beta_{x,+}^* = \beta_{x,-}^* = \beta_{y,+}^* = \beta_{y,-}^* \equiv \beta^*$ , $N_+ = N_- \equiv N$ ,<br>$E_+ = E_- \equiv E$ , $D_{x,\pm}^* = D_{y,\pm}^* = 0$                          |
| $\frac{\pi f_c \gamma^2 \epsilon_x \xi_x \xi_y (1+r)^2}{r_e^2 \beta_y^*}$  | $\epsilon_{x,+} = \epsilon_{x,-} \equiv \epsilon_x$ , $\sigma_{y,+}^* = \sigma_{y,-}^*$ ,<br>$\beta_{x,+}^* = \beta_{x,-}^* \equiv \beta_x^*$ , $\beta_{y,+}^* = \beta_{y,-}^* \equiv \beta_y^*$ ,<br>$N_+ = N_- \equiv N$ , $E_+ = E_- \equiv E$ , $D_{x,\pm}^* = D_{y,\pm}^* = 0$ |

**Alternative expressions** Because the luminosity in a circular collider is limited by the value of  $\xi$ , it is useful to write  $\mathcal{L}$  explicitly in terms of  $\xi$  as seen in the third row of Tab.1. Here  $E$  and  $I$  are the beam energy and total beam current in one ring and  $K = 1/(2e^3) = 1/(2er_e m_e c^2)$ . With  $E$  in GeV,  $I$  in A,  $\beta_y^*$  in cm, and  $\mathcal{L}$  in  $\text{cm}^{-2}\text{s}^{-1}$ , we have  $K = 2.17 \times 10^{34}$ . The symbol  $( )_{+,-}$  means that the enclosed parameters may be taken from either beam, on account of the transparency conditions.

For a linear collider,

$$\mathcal{L} = \frac{H}{4\pi E} \frac{N}{\sigma_x^*} \frac{P}{\sigma_y^*} \quad (4)$$

where  $N = N_+ = N_-$ ,  $P$  is the average beam power, and  $H$  is the pinch enhancement factor [12]. The factor  $N/\sigma_x^*$  determines the number of beamsstrahlung photons emitted (constrained by background considerations). The factor  $P/\sigma_y^*$  represents the major technical challenge—providing high beam power and small bunch size.

**Reductions and enhancements of the luminosity** When  $\sigma_z \gtrsim \beta^*$ , the luminosity decreases relative to the case when  $\sigma_z \ll \beta^*$  due to the reduced overlap of the two bunches during the collision. This geometrical (“hourglass”) reduction

factor for Gaussian beams is [13]

$$R(t_x, t_y) \equiv \frac{\mathcal{L}}{\mathcal{L}_0} = \int_{-\infty}^{\infty} \frac{dt}{\sqrt{\pi}} \frac{\exp(-t^2)}{\sqrt{(1+t^2/t_x^2)(1+t^2/t_y^2)}} \quad (5)$$

with

$$t_x^2 = \frac{2(\sigma_{x,+}^{*2} + \sigma_{x,-}^{*2})}{(\sigma_{z,+}^2 + \sigma_{z,-}^2)(\sigma_{x,+}^{*2}/\beta_{x,+}^{*2} + \sigma_{x,-}^{*2}/\beta_{x,-}^{*2})}$$

and correspondingly for  $t_y$ . The nominal luminosity,  $\mathcal{L}_0$ , is that represented by Tab.1, which corresponds to the limit  $\sigma_z \rightarrow 0$ . See Fig.1 [13].

Another reduction factor comes from a non-zero crossing angle, typically in the horizontal plane. For the symmetric-collider case with  $\sigma_y^* \ll \sigma_x^*$  we obtain [14]

$$R_L \equiv \frac{\mathcal{L}}{\mathcal{L}_0} = \sqrt{\frac{2}{\pi}} a e^b K_0(b), \quad (6a)$$

$$a = \frac{\beta_y^*}{\sqrt{2}\sigma_z}, \quad (6b)$$

$$b = a^2 \left[ 1 + \left( \frac{\sigma_z}{\sigma_x^*} \tan \phi \right)^2 \right] \quad (6c)$$

where  $K_0$  is a Bessel function and  $\phi$  is half the crossing angle. For the symmetric-collider case

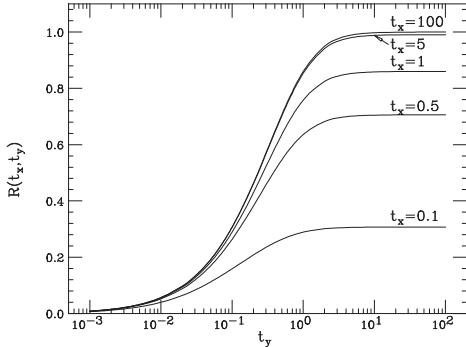


Figure 1: Hourglass reduction factors, Eq.(5).

with round beams  $\sigma_y^* = \sigma_x^*$  and  $\beta_x^* = \beta_y^* \equiv \beta^*$  (typical for hadron colliders) one has [15]

$$R_L \equiv \frac{\mathcal{L}}{\mathcal{L}_0} = \frac{\cos \phi}{\sqrt{\pi} \sigma_z} \int_{-\infty}^{\infty} \frac{e^{-As^2}}{1 + (s/\beta^*)^2} ds , \quad (7a)$$

$$A \equiv \frac{\sin^2 \phi}{\sigma_x^{*2} (1 + (s/\beta^*)^2)} + \frac{\cos^2 \phi}{\sigma_z^2} . \quad (7b)$$

When  $\sigma_z \ll \beta_y^*$ , Eqs.(6) and (7) reduce to [4]

$$R_L = \left[ 1 + \left( \frac{\sigma_z}{\sigma_x^*} \tan \phi \right)^2 \right]^{-1/2} \quad (8)$$

If the beams are, in addition, offset transversely by  $\delta x$  and  $\delta y$ , Eq.(8) acquires the extra multiplicative factor

$$\exp \left\{ - \frac{(\delta x/2)^2}{\sigma_x^{*2} \cos^2 \phi + \sigma_z^2 \sin^2 \phi} - \left( \frac{\delta y}{2\sigma_y^*} \right)^2 \right\}$$

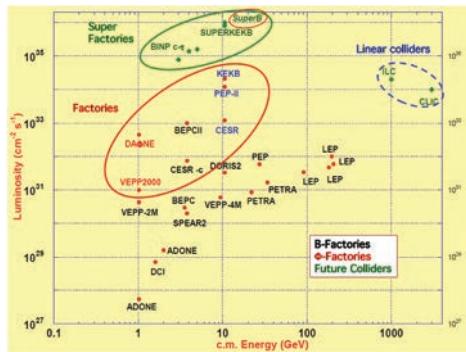
**Monochromatization** If the purpose of the collider is the study of a very narrow resonance, i.e. narrow compared to the beam energy spread, such as a  $\tau$ -charm factory, it is possible to enhance the luminosity by means of a monochromatization scheme. For example, the  $J/\psi$  resonance has a width-to-mass ratio  $\Gamma/mc^2 = 2.8 \times 10^{-5}$ , small compared to the energy spread of the beam, typically  $\sigma_\delta \lesssim 10^{-3}$ . The improved energy resolution also allows detailed measurement of thresholds and branching rates in the  $e^+e^-$  c.m. energy range 3 – 5 GeV [16, 17]. The scheme calls for large  $y$  dispersion such that  $D_{y,+}^* = -D_{y,-}^*$ , which induces a correlation between the  $y$  position of the particles and their energy in such a way that  $e^+$ 's with higher-than-average energy preferentially collide with  $e^-$ 's with lower-than-average

energy and vice versa. Thus the energy spread of the c.m. of any given colliding  $e^+e^-$  pair is effectively reduced relative to the standard (zero dispersion) case, hence the name “monochromatization” [18].

**Crab crossing scheme** The simplest design for the interaction region (IR) of all particle colliders implies that the beams cross at an angle, hence the luminosity is subject to the reduction factor  $R_L$ , Eq.(8), relative to the idealized case  $\phi = 0$ . The PEP-II design [6] was perhaps the only exception, achieving  $\phi = 0$  by means of a rather unusual IR geometry that involved additional vertical and horizontal bends in the positron beam. In the more generic case, however, reducing  $\phi$  implies long IRs, which may only be practical for very large machines such as hadron colliders. Even then, if the beam is made up of many closely-spaced bunches, as is the case in the design of all present and future colliders, a small value of  $\phi$  has the drawback of giving rise to a number of undesirable “parasitic,” or “long-range,” collisions between the bunches in the opposing beams as they travel towards the IP in a common vacuum pipe. A way to compensate for  $\phi$  is to tilt the bunches by an angle  $\phi$  such that they will collide with zero effective crossing angle. Such tilting can be achieved by applying a time-dependent transverse dipolar kick to the bunches as they approach the IP [19, 20, 21]. See Sec.4.13.

**Crab waist scheme** An alternative to the crab-crossing scheme has been developed and successfully tested at the DAΦNE  $e^+e^-$  collider. In this scheme the bunches collide in the  $x - z$  plane ( $x$ =horizontal dimension,  $z$ =longitudinal). The bunches are tilted so that  $\phi$  is relatively large, but the collision is arranged such that the longitudinal location of the vertical waist is a function of  $x$ . Here all components of the bunch collide at the minimum of  $\beta_y$ , hence the hourglass reduction is lessened, and so are detrimental beam-beam dynamical effects [22, 23, 24]. See Sec.4.13.

**Peak and average luminosity** The history of peak luminosity achieved or designed for many colliders is illustrated in Figs.2 and 3. Roughly speaking, the typical peak luminosity for  $e^+e^-$  colliders is 1–2 orders of magnitude larger than for hadron colliders. This difference can be attributed to several factors: leptons, being point-like objects, tend to yield cleaner reactions upon collision than hadrons, which are composite.

Figure 2: Peak luminosity of  $e^+e^-$  colliders.

Therefore, leptons can be collided at a lower energy than hadrons to study specific physical processes. As a result, lepton colliders tend to be smaller, hence less costly, than hadron colliders for a given physics-equivalent beam energy. This advantage of lepton colliders diminishes as the beam energy increases owing to the much stronger emission of synchrotron radiation than hadrons, related to the relative lightness of the leptons. Thus, the LHC has a much wider physics reach than LEP, which was of the same size as the LHC, owing to the overwhelming effect from synchrotron radiation at LEP. As a positive side effect of the higher synchrotron radiation, lepton beams are intrinsically more stable than hadron beams, hence lepton colliders can typically accommodate higher beam intensities than their hadron counterparts. In addition, electron beams are typically less expensive to produce than hadron beams.

While peak luminosity is typically a basic performance parameter specified a priori in the collider design, it is the average luminosity, or its integral over a period of time, that ultimately defines the success of the detailed physics obtained from the collider. To account for down time, injection, beam lifetimes, etc., one experimental “year” is taken by convention to be  $10^7$  s. Then, the expected integrated luminosity for a collider delivering a peak luminosity of  $1 \times 10^{33} \text{ cm}^{-2} \text{ s}^{-1}$  would be  $\int \mathcal{L} dt = 1 \times 10^{40} \text{ cm}^{-2} = 10 \text{ fb}^{-1} = 10,000 \text{ pb}^{-1}$  ( $1 \text{ b} = 10^{-24} \text{ cm}^2$ ). In practice, however, the long-term integrated luminosity depends on many more variables than peak luminosity, and many of them are unrelated to beam dynamics. Perhaps the simplest variable that can be controlled is the duration of a “store”  $t_c$  (i.e., the

length of time during which the beams are in collision). Following injection, the luminosity decays in time due to particle losses from various sources. If it takes a time  $t_f$  to refill the machine and bring the beams back into collision, during which time the beams are not colliding, one wants to determine the value of  $t_c$  that leads to the largest average luminosity. If we make the approximation  $\mathcal{L}(t) = \mathcal{L}_0 \exp(-t/\tau)$  where  $\tau$  is the characteristic lifetime, then the average luminosity is given by<sup>2</sup>

$$\langle \mathcal{L} \rangle = \frac{1}{t_c + t_f} \int_0^{t_c} dt \mathcal{L}(t) = \mathcal{L}_0 \frac{1 - e^{-x}}{x + a} \quad (9)$$

where  $x = t_c/\tau$  and  $a = t_f/\tau$ . If  $t_f$  is independent of the number of particles left in the machine at the end of the luminosity run, the equation for  $t_c$  (or, equivalently,  $x$ ) that maximizes  $\langle \mathcal{L} \rangle$  that follows from (9) is [25]

$$e^x = 1 + x + a \quad (10)$$

Given  $a$ , Eq.(10) can be readily solved numerically by iterating the map  $x_{n+1} = \ln(1 + x_n + a)$  starting from any  $x_0 > -a$ . An approximate solution is given by the first iteration,  $x_1$ , assuming the choice  $x_0 = \sqrt{2a}$ ,

$$x \simeq \ln \left( 1 + \sqrt{2a} + a \right)$$

This expression becomes exact when  $a \rightarrow 0$ . For larger  $a$ , its relative error is at most  $\sim 7\%$ , and this worst case occurs for  $a \simeq 1.1$ . Thus, if the condition (10) is satisfied, the maximum average luminosity is

$$\langle \mathcal{L} \rangle_{\max} = \mathcal{L}_0 e^{-x} \simeq \frac{\mathcal{L}_0}{1 + \sqrt{2a} + a} \quad (11)$$

If  $t_f$  does depend on the number of particles left in the machine at the end of the prior luminosity run, the optimal condition is, of course, more complicated [25], although a similar analysis is applicable.

Figs.4, 5, 6 and 7 show the luminosity history of CESR, PEP-II, KEKB and DAΦNE, respectively; Figs.8, 9 and 10 give the luminosity history of the Tevatron and RHIC; and Figs.11, 12 and 13 the luminosity history of the LHC during 2011.

The time evolution of the luminosity performance of many past and present colliders exhibits systematic features that can be interpreted in terms of the complexity of the machine [27].

<sup>2</sup>The exponential decay is a convenient approximation; for a more detailed analysis see M.S. Zisman, Sec.3.3.2.2.

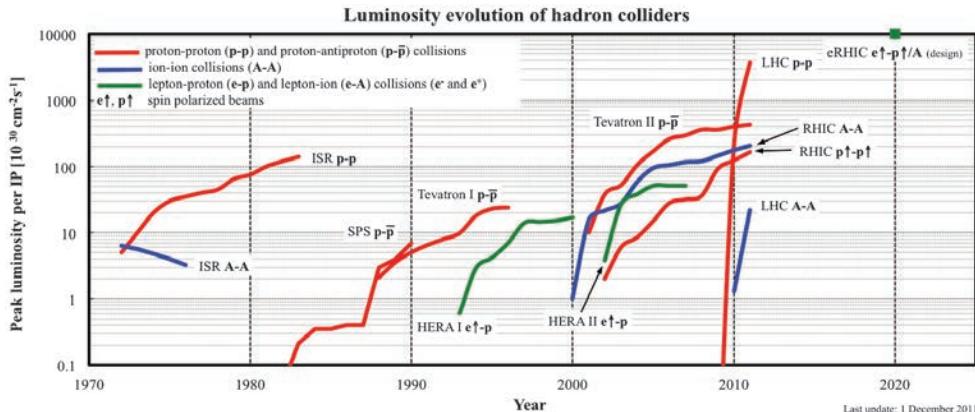


Figure 3: Peak luminosity of hadron and  $e$ - $p$  colliders. Arrows indicate polarized beams.

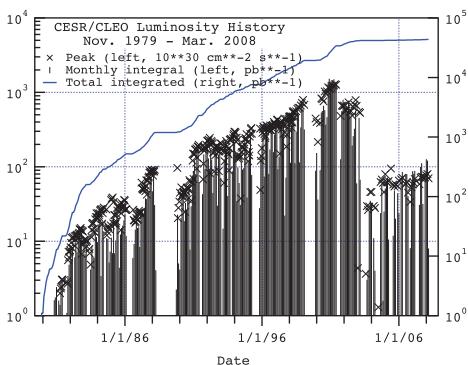


Figure 4: Luminosity history of CESR. The lower luminosity after mid-2003 was due to a reduction in beam energy from 5.3 to 1.9 GeV, which was undertaken to run CESR in charm-factory mode [26].

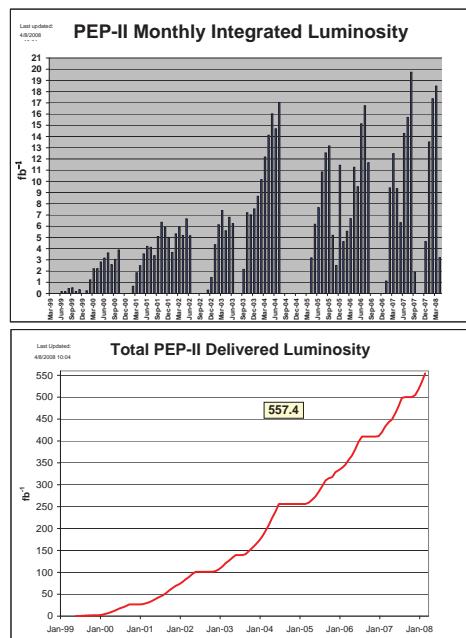


Figure 5: History of PEP-II luminosity. Top: monthly integral; bottom: total integrated.

- ## References
- [1] We are grateful to H. Burkhardt, W. Fischer, O. Nopoly, F. Porter, A. Zholents and F. Zimmermann for valuable comments. We are indebted to D. Rice, N. Phinney, J. Marriner, V. Shiltsev, R. Moore, Y. Funakoshi and S. Guiducci for providing the luminosity data.
  - [2] C. Møller, K. Danske Vidensk. Selsk. Mat.-Fys. Medd. 23 (1945) 1
  - [3] M. Furman, LBNL-53553 (2003)
  - [4] O. Nopoly, PA 40 (1993) 180
  - [5] D. Edwards, M. Syphers, An Introduction to the Physics of High-Energy Accelerators, John Wiley & Sons, NY, 1993, p.9
  - [6] “PEP-II: An Asymmetric B Factory – Conceptual Design Report,” LBL-PUB-5379, SLAC-418, (1993) p.184
  - [7] “KEKB B-Factory Design Report,” KEK Report 95-7
  - [8] A.A. Garren et al, PAC 89, 1847

### Sec.4.1: LUMINOSITY

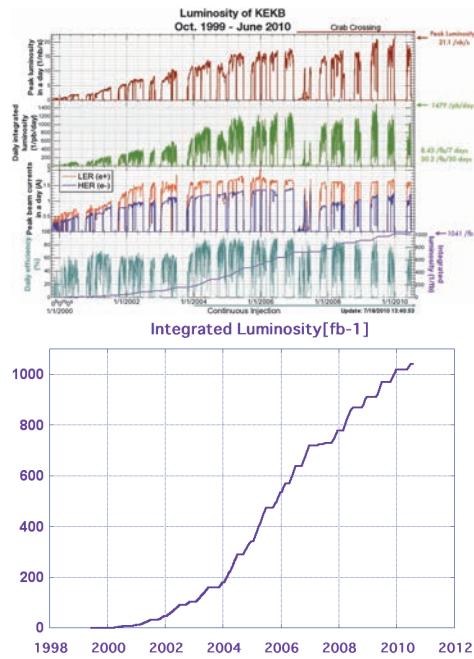


Figure 6: KEKB luminosity history. Top: peak luminosity and other quantities, as indicated. Bottom: total integrated.

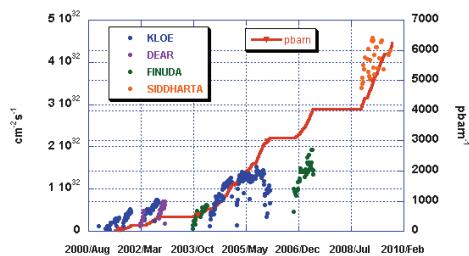


Figure 7: DAΦNE luminosity history. Dots indicate peak luminosity at the various detectors employed (left-hand scale). The solid line is the total integrated luminosity (right-hand scale). The increase in luminosity delivered to SIDDHARTA was achieved with the crab-waist scheme (see text).

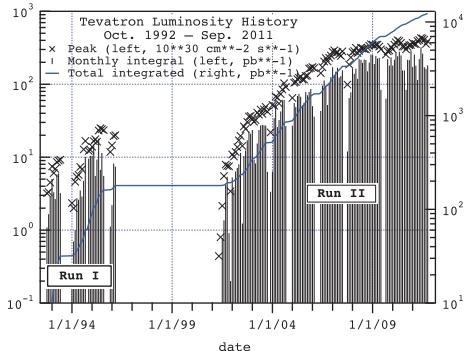


Figure 8: Luminosity history of the Tevatron. The 4-year shutdown between Runs I and II was undertaken to carry out improvements that included: (a) construction of a new synchrotron (the Main Injector) to increase the proton bunch intensity in the Tevatron, and the proton flux on the  $\bar{p}$  target; (b) construction of a new  $\bar{p}$  storage ring (the Recycler), originally intended to store leftover  $\bar{p}$ 's at the end of collider stores; (c) a variety of improvements to the existing Antiproton Source, namely doubling the bandwidth of the stochastic cooling systems; changing the transition  $\gamma$  in the Antiproton Accumulator; and moving the injection location in the Tevatron in order to accept beams from the Main Injector [28].

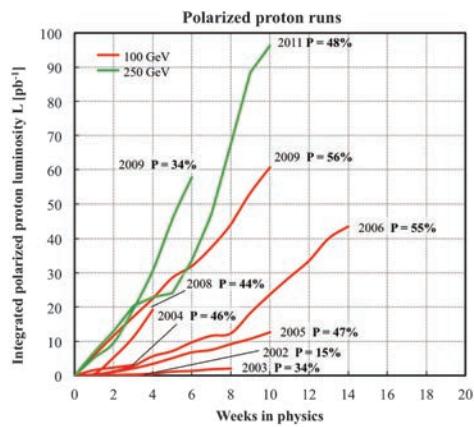


Figure 9: RHIC luminosity for polarized proton beams.

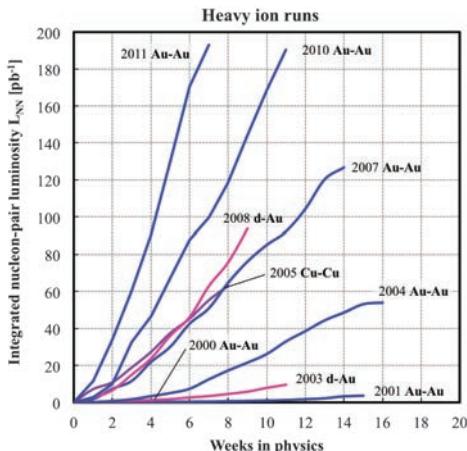


Figure 10: RHIC luminosity for ion beams. The “ion-pair luminosity”  $\mathcal{L}_{N_1 N_2}$  is defined to be  $N_1 N_2 \mathcal{L}$ , where  $N_i$  ( $i = 1, 2$ ) is the number of nucleons per ion in beam  $i$ , and  $\mathcal{L}$  is the conventional luminosity for ion-pair collisions.

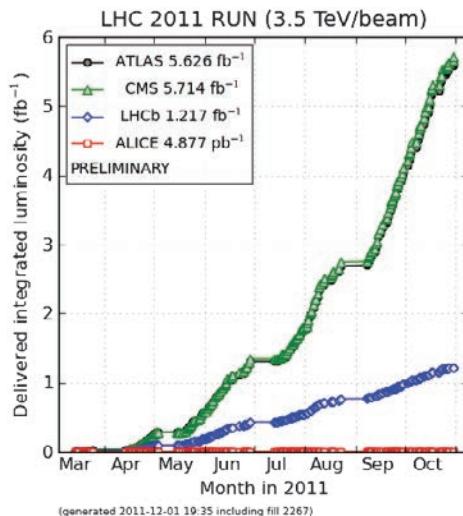


Figure 12: LHC integrated luminosity delivered to the four detectors during 2011, for p–p collisions.

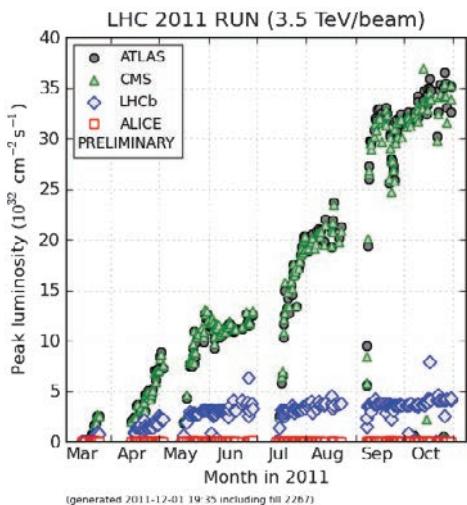


Figure 11: LHC peak luminosity for p–p collisions as observed during 2011 at the four detectors.

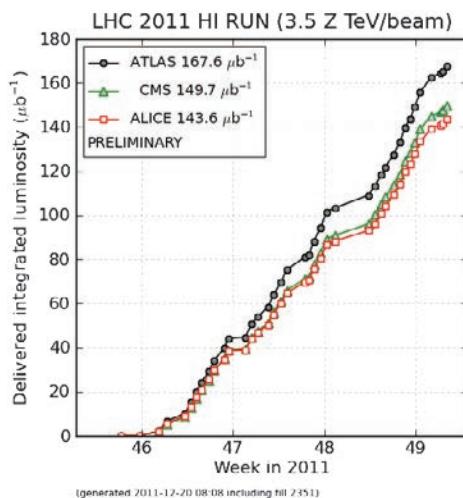


Figure 13: LHC integrated luminosity delivered to three of the four detectors during 2011, for  $\text{Pb}^{82+}$ – $\text{Pb}^{82+}$  collisions.

## Sec.4.2: BRIGHTNESS

- [9] Y.H. Chin, AIP Conf. Proc. 214 (1990) p.424
- [10] S. Krishnagopal, R. Siemann, PR D41 (1990) 1741
- [11] Y. Funakoshi, ICFA Beam Dynamics Newsletter No.31 (2003) p.78
- [12] K. Yokoya, P. Chen, US-CERN School on Particle Accelerators, Frontiers of Particle Beams: Intensity Limitations (Hilton Head Island, S. Carolina, USA, Nov. 714, 1990) Springer Verlag LNP 400, p.415
- [13] M.A. Furman, PAC 91, 422
- [14] K. Hirata, PRL 74 (1995) 2228
- [15] B. Muratori, CERN AB-Note-2003-026 (ABP)
- [16] J. Jowett, Springer-Verlag Lecture Notes in Phys. 425 (1992), p.79
- [17] J. Kirkby, AIP Conf. Proc. 349, p.11
- [18] A. Renieri, Frascati preprint INF-75/6(R), 1975
- [19] R. Palmer, SLAC-PUB-4707 (1988), unpublished.
- [20] K. Oide, K. Yokoya, PR A40 (1989) 315
- [21] K. Ohmi, ICFA Beam Dynamics Newsletter 52 (2010) p.33
- [22] P. Raimondi, 2nd SuperB Workshop, Frascati, Mar. 2006,  
<http://www.lnf.infn.it/conference/superb06/talks/raimondi1.ppt>
- [23] P. Raimondi et al, LNF-07/003; arXiv: physics/0702033
- [24] M. Zobov et al, PRL 104, 174801 (2010)
- [25] F.C. Porter, NIM A302 (1991) 209
- [26] D. Rice, D. Rubin, ICFA Beam Dynamics Newsletter No. 48 (2009) p.152
- [27] V. Shiltsev, Mod. Phys. Lett. A 26, 11 (2011) 761
- [28] S. Holmes et al, JINST 6, T08001 (2011)

## 4.2 BRIGHTNESS

*P. Elleaume, Deceased 2011  
 K.-J. Kim, ANL*

Particle density in phase space is generally referred to as the brightness distribution. The brightness distribution plays an important role in beam transport calculation and an invariant characterization of the source strength. The concept applies to both particle and photon beams.

### 4.2.1 Particle Beam

**Brightness distribution** The brightness distribution  $\mathcal{B}$  is the density in 6-D phase space [1]  $(x, p_x, y, p_y, t, E)$  where  $t$  is the arrival time,  $E$  is the kinetic energy canonical conjugate to  $t$ . When the system has a Hamiltonian,  $\mathcal{B}$  is invariant along each particle trajectory in an accelerator.

When  $\gamma \gg 1$ , the phase space coordinates  $\approx (x, \gamma x', y, \gamma y', z, \Delta\gamma)$ , where  $z$  is the particle position relative to beam center, and  $\Delta\gamma = \gamma - \gamma_0$  with  $\gamma_0$  energy of the reference particle. Without acceleration, another convenient set of the phase space coordinates is  $(x, x', y, y', z, \Delta\gamma)$ .  $\mathcal{B}$  is defined as the density in the appropriate phase space.

Assume  $\gamma \gg 1$  with no acceleration. Assume  $\mathcal{B}$  factorizes in the three dimensions, and consider  $x$ -dimension. (Extension to general case is straightforward.)

$$\mathcal{B}(x, x'; s) = \frac{d^2 F}{dx dx'} \quad (1)$$

where  $F$  may be considered as the flux or longitudinal particle density. The  $\mathcal{B}$  distributions at two different  $s$  are related by the coordinate transformation between them,

$$\mathcal{B}(x_2, x'_2; s_2) = \mathcal{B}(x_1, x'_1; s_1) \quad (2)$$

$$\begin{pmatrix} x_1 \\ x'_1 \end{pmatrix} = \mathbf{M}^{-1} \begin{pmatrix} x_2 \\ x'_2 \end{pmatrix}$$

The spatial and the angular densities of the flux are

$$S(x; s) = \frac{dF}{dx} = \int B(x, x'; s) dx' \quad (3)$$

$$A(x'; s) = \frac{dF}{dx'} = \int B(x, x'; s) dx \quad (4)$$

$$F = \int S(x; s) dx = \int A(x'; s) dx' = \int B(x, x') dx dx'$$

$S$  and  $A$  are not invariant along the particle trajectory. In the absence of aperture,  $F$  is conserved and is an invariant characterization of the global strength of the beam.

**Brightness** For a well-designed beam,  $\mathcal{B}$  is a smooth function peaked at the phase space origin. Thus  $\mathcal{B}(\text{origin})$  is often referred to as the brightness. A related quantity is the emittance (phase space area). The brightness is flux divided by emittance. There are different definitions of the emittance, and hence of the brightness. One definition is the rms emittance [2]  $\epsilon_x = \sqrt{\langle x^2 \rangle \langle x'^2 \rangle - \langle xx' \rangle^2}$  where  $\langle \rangle$  means averaging with  $\mathcal{B}$  as the weight function.

Near the beam waist at  $s = 0$ , one may have a Gaussian brightness distribution

$$\mathcal{B}(x, x'; 0) = \frac{F}{2\pi\epsilon_x} \exp\left[-\frac{1}{2} \left( \frac{x^2}{\sigma_x^2} + \frac{x'^2}{\sigma_{x'}^2} \right)\right] \quad (5)$$

with brightness  $\mathcal{B}_0 = F/(2\pi\epsilon_x)$ . See also Sec.3.1.2.

**Emittance** The rms emittance is invariant for relativistic beam with no acceleration. With acceleration, the normalized emittance  $\epsilon_{Nx} = \gamma\beta_x$  is invariant (Sec.2.2.1).

For applications involving bunch compression, consider longitudinal brightness distribution [3]

$$\mathcal{B}_L(z, \Delta\gamma; s) = \frac{d^2N}{dz d\Delta\gamma} \quad (6)$$

which in the relativistic limit is invariant under beam compression and acceleration. For a Gaussian distribution, the peak longitudinal brightness is  $\mathcal{B}_L = N/(2\pi\sigma_z\sigma_{\Delta\gamma})$  and  $\epsilon_z = \sigma_z\sigma_{\Delta\gamma}$  the normalized longitudinal emittance.

## References

- [1] E.D. Courant, H.S. Snyder, Ann. Phys. 3 (1958) 1
- [2] P.M. Lapostolle, CERN-ISR-DI/71-6 (1971)
- [3] C. Pellegrini, NIM A272 (1988) 364

## 4.2.2 Radiation Beam

### 4.2.2.1 Introduction

The concepts in Sec.4.2.1 also apply for the radiation beams in geometric optics. In wave optics, the concept of phase space and brightness do not have direct physical meaning. Nevertheless, it is possible to apply these concepts by identifying the Wigner Distribution of the electric fields as the brightness distribution [4]. For a Gaussian mode (TEM00) propagating in free space, Eq.(5) holds. The rms radiation size and angular divergence are  $\sigma_R \equiv \sigma_x = \sigma_y$  and  $\sigma'_R \equiv \sigma'_x = \sigma'_y$  (cylindrical symmetry). For a coherent Gaussian mode , the “radiation rms emittance”  $\sigma'_R\sigma_R$  is equal to  $\lambda/4\pi$ . The quantity corresponding to the beta function in radiation optics is called the Rayleigh length  $z_R = \sigma_R/\sigma'_R$ . See also Sec.3.1.2. In general, the brightness distribution of synchrotron radiation is the convolution of single electron brightness (also called diffractive brightness) with the electron brightness distribution. The radiation beam brightness for synchrotron radiation is usually expressed in units of photons/sec/mm<sup>2</sup>/mr<sup>2</sup>/0.1% spectral bandwidth. In the literature, there exist several denominations such as brightness, spectral brightness, brilliance or spectral brilliance which all refer to the same quantity. Radiation brightness is in general a function of the transverse phase space ( $x, x', y, y'$ ) and photon energy  $\varepsilon$ . The brightness of radiation with specific

polarization state can be defined by introducing four brightness functions in analogy with the four Stokes parameters (Sec.3.1.5.1) defining the polarization of incoherent photon beams [5]. In most cases of interest, the spectral brightness is maximum on axis of the electron beam ( $x, x', y, y' = (0, 0, 0, 0)$ ). In the following we shall restrict the discussion to on-axis brightness which only depends on  $\varepsilon$  and in most cases corresponds to linear polarization with horizontal electric field.

### 4.2.2.2 Bending magnet radiation

We consider an electron beam circulating along an horizontal planar circular trajectory in a bending magnet with vertical magnetic field. The on axis brightness  $B_{BM}$  of the synchrotron radiation produced in such a bending magnet is expressed as :

$$B_{BM} = \frac{d\Phi}{d\Omega d\varepsilon/\varepsilon} \frac{1}{2\pi\sqrt{\sigma_x^2 + \sigma_R^2}\sqrt{\sigma_y^2 + \sigma_R^2}} \quad (1)$$

Where  $\sigma_x$  ( $\sigma_y$ ) is the rms horizontal (vertical) electron beam size at the source point and  $\frac{d\Phi}{d\Omega d\varepsilon/\varepsilon}$  is the angular spectral flux generated per unit solid angle  $d\Omega$  in the plane of the orbit per unit relative energy bandwidth  $d\varepsilon/\varepsilon$  of the photons.  $\frac{d\Phi}{d\Omega d\varepsilon/\varepsilon}$  is expressed as a function of the electron energy  $E$ , the electron current  $I$ , the bending magnet field  $B$  and photon energy  $\varepsilon$  according to [6]:

$$\frac{d\Phi}{d\Omega d\varepsilon/\varepsilon} [\text{photons/s/mr}^2/0.1\%] = \quad (2)$$

$$1.327 \times 10^{13} E^2 [\text{GeV}] I[\text{A}] \frac{\left(\frac{\varepsilon}{\varepsilon_c}\right)^2 K_{2/3}^2 \left(\frac{\varepsilon}{\varepsilon_c}\right)}{\sqrt{1 + \left(\frac{\sigma'_y}{\sigma'_R}\right)^2}}$$

where  $K_{2/3}(y)$  is modified Bessel function of order 2/3,  $\varepsilon_c$  the critical energy of the bending magnet radiation,  $\varepsilon_c [\text{keV}] = 0.665E^2 [\text{GeV}] B [\text{T}]$ ,  $\sigma'_y$  the vertical rms divergence of the electron beam and  $\sigma'_R(\sigma_R)$  the natural rms divergence (rms size) produced by a single electron (diffractive contributions), which depends on photon energy  $\varepsilon = hc/\lambda$ , and is approximated as [7]

$$\sigma'_R [\mu\text{rad}] = \frac{320}{E [\text{GeV}]} \left(\frac{\varepsilon}{\varepsilon_c}\right)^{-x} \quad (3)$$

and  $\sigma'_R\sigma_R \cong \lambda/(4\pi)$ . with  $x = 0.354$  (0.549) if  $\varepsilon/\varepsilon_c < 1(> 1)$ ,  $h$  the Planck constant and  $c$  the speed of light.

### 4.2.2.3 Wiggle radiation

In a planar sinusoidal vertical field undulator or wiggler, the electron beam performs an oscillating motion characterized by a deflection parameter  $K$  which is related to the spatial period  $\lambda_0$  of the magnetic field and the peak magnetic field  $\hat{B}$  according to  $K = 0.0934\lambda_0[\text{mm}]\hat{B}[\text{T}]$ . One differentiates two limiting cases namely the undulator case ( $K < 2.5$ ) in which the radiation from each period interferes producing an harmonic spectrum and the wiggler case ( $K > 3$ ) for which the radiation originating from each period has a nearly random phase justifying the derivation of brightness as the simple sum of brightness from each source point.

The on axis brightness  $B_W$  of the wiggler radiation is [6]

$$B_W = \frac{d\Phi}{d\Omega d\varepsilon/\varepsilon} \frac{2N}{2\pi\Sigma_x\Sigma_y} \quad (4)$$

where  $\frac{d\Phi}{d\Omega d\varepsilon/\varepsilon}$  is the angular spectral flux generated on axis by a bending magnet of field  $\hat{B}$  given by Eq.(2), and  $N$  the number of periods of the magnetic field.  $\Sigma_x$  and  $\Sigma_y$  are the effective horizontal and vertical rms source sizes averaged over the length  $L$  of the wiggler

$$\begin{aligned} \Sigma_x^2 &= \sigma_x^2 + a^2 + \frac{L^2}{12}\sigma_x'^2 + \sigma_R^2 \quad \text{and} \\ \Sigma_y^2 &= \sigma_y^2 + \frac{L^2}{12}\sigma_y'^2 + \sigma_R^2 \end{aligned} \quad (5)$$

with  $\sigma_R$  defined by Eq.(3) and  $a$  the amplitude of the oscillation of the horizontal trajectory inside the wiggler, which is given by :

$$a[\mu\text{m}] = \frac{\lambda_0[\text{mm}]K}{2\pi 1.957 E[\text{GeV}]} \quad (6)$$

### 4.2.2.4 Undulator radiation

The spectrum of the radiation produced by an undulator is made of an harmonic series of peaks.

For the most commonly used planar field undulator, the energy  $\varepsilon_n$  of these peaks observed on axis of the trajectory, is given by:

$$\varepsilon_n[\text{keV}] = 9.50 \frac{nE^2[\text{GeV}]}{\lambda_0[\text{mm}] \left(1 + \frac{K^2}{2}\right)} \quad (7)$$

Where  $n$  is an integer corresponding to the harmonic number. The Brightness  $B_U$  of undulator radiation is maximum on axis of the trajectory for

the most intense odd harmonics of the spectrum, it is approximated by [8]:

$$B_U = \frac{\Phi_n}{(2\pi)^2 \Sigma_x \Sigma_y \Sigma'_x \Sigma'_y} \quad (8)$$

Where  $\Phi_n$  is the total spectral flux generated on the nth harmonic which depends on the electron current  $I$ , number of period  $N$  of the magnetic field and deflection parameter  $K$ :

$$\begin{aligned} \Phi_n &[\text{Photons/s/0.1\%}] \quad (9) \\ &= 1.431 \times 10^{14} NI[A] \frac{nK^2}{1 + \frac{K^2}{2}} \\ &\quad \left[ J_{\frac{n-1}{2}} \left( \frac{nK^2}{4 + 2K^2} \right) - J_{\frac{n+1}{2}} \left( \frac{nK^2}{4 + 2K^2} \right) \right]^2 \end{aligned}$$

and  $\Sigma_x = \sqrt{\sigma_x^2 + \sigma_R^2}$  and  $\Sigma'_x = \sqrt{\sigma_x'^2 + \sigma_R'^2}$ ; similarly for the y direction. The size  $\sigma_R$  and divergence  $\sigma'_R$  correspond to single electron emission (diffractive contribution) with  $\sigma_R \cong \sqrt{\frac{\lambda}{2L}}$  and  $\sigma'_R \cong \sqrt{\frac{\lambda L}{8\pi^2}}$ .

As can be seen from Eq.(8), the undulator brightness increases as the beam size and divergence of the electron beam shrinks. Since the brightness is an important figure of merit, synchrotron light sources are optimized for a small emittance of the electron beam. A facility incorporating many undulators in a storage ring designed for high radiation brightness featuring small emittance, high current, and long straight sections is referred to as “3rd generation” facility.

### 4.2.2.5 Brightness comparison

Fig.1 compares on axis brightness from a bending magnet, a wiggler and an undulator installed in the ESRF storage ring. Note that the brightness from the wiggler is computed for the maximum field of 1.5 T. On the other hand the brightness curves for the undulator require to adjust the field to the targeted photon energy defined in Eq.(7). Each harmonic covers a different photon energy range. Note that there exists a lack of overlap of the spectral range covered by harmonics 1 and 3 which is linked to the low value ( $K = 1.8$ ) of the U23 undulator. High  $K$  undulators provide increased tunability at the cost of larger integrated angle, lower energy of the fundamental peak and higher harmonic content. On the other hand low  $K$  undulators provide limited tunability

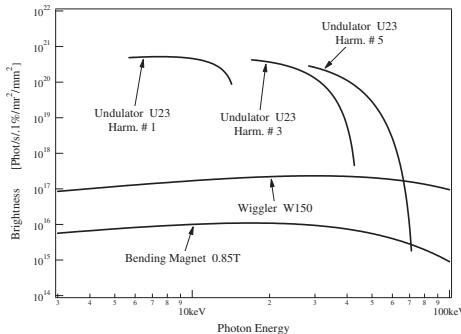


Figure 1: Comparison of brightness from several sources of radiation installed at the ESRF. The electron beam energy is 6 GeV, the ring current 200 mA, the horizontal and vertical emittances are 4 nm and 10 pm. The U23 undulator has 174 periods of 23 mm and a peak field of 0.83 T. The wiggler W150 has a 150 mm period and a length of 1.5 m.

and are useful only on the lowest harmonic numbers. Note that some brightness can also be obtained on the even harmonics of the spectrum but their value is usually one or two orders of magnitude lower and it is difficult to derive an analytical estimate.

The brightness from other third generation is similar but in general shifted to lower photon energies for lower electron energy facilities (ALS, SLS,...) and to higher photon energies for higher electron energy facilities (APS, SPring-8). The emerging “4th” generation facilities employing free electron lasers have much higher peak brightness. Detailed information on the source characteristics of these facilities can be found in [9].

## References

- [1] E.D. Courant, H.S. Snyder, Ann.Phys. 3 (1958) 1
- [2] P.M. Lapostolle, CERN-ISR-DI/71-6 (1971)
- [3] C. Pellegrini, NIM A272 (1988) 364
- [4] K.-J. Kim, NIM A246 (1986) 71
- [5] M. Born, E. Wolf, Principles of Optics, Pergamon Press (1986)
- [6] Undulators, Wigglers and their Applications, Edited by H. Onuki & P. Elleaume, Taylor and Francis (2003)
- [7] K.-J. Kim, X-ray Data Booklet, Center for X-ray Optics and Advanced Light Source, PUB-490 rev 2 (<http://xdb.lbl.gov/>)
- [8] K.-J. Kim, AIP Proc. 184 (1989) 565
- [9] <http://www.lightsources.org>

## 4.3 LINAC OPERATION

### 4.3.1 Operation of High Energy Electron Linacs

*T.O. Raubenheimer, SLAC*

(See Refs.[1, 2, 3].)

**Linac model** The longitudinal equation of motion in a long linac is

$$\frac{d}{ds}\gamma(s, z) = \frac{e}{mc^2} \mathcal{E}_{\text{rf}}(s) \cos(k_{\text{rf}}z + \phi_{\text{rf}}) + 4\pi\epsilon_0 N_B r_e \int_z^\infty W_{\parallel}(s, z' - z)\lambda(z')dz'$$

where  $s, z$  = longitudinal position in the accelerator and in the bunch,  $\mathcal{E}_{\text{rf}}$  = acceleration gradient, and  $k_{\text{rf}}$  = rf wavenumber,  $\phi_{\text{rf}}$  = rf phase at the bunch center,  $W_{\parallel}$  = longitudinal wakefunction (V/C/m),  $\lambda$  = longitudinal beam distribution.

The transverse equation of motion is

$$\begin{aligned} \frac{1}{\gamma} \frac{d}{ds}\gamma \frac{d}{ds}y(s, z) - (1 - \delta)K(y - y_q) \\ = (1 - \delta)\frac{1}{\rho} - 4\pi\epsilon_0 N_B r_e \frac{1 - \delta}{\gamma_0} \\ \times \int_z^\infty W_{\perp}(s, z' - z)(y(z) - y_a)\lambda(z')dz' \end{aligned} \quad (1)$$

where  $\delta(s, z) = (\gamma - \gamma_0)/\gamma$ ,  $\gamma_0 mc^2$  = nominal beam energy,  $K$  = normalized quadrupole field,  $\rho$  = bending radius,  $y_q, y_a$  = quadrupole and accelerator structure misalignments, and  $W_{\perp}$  = transverse wakefunction (V/C/m²).

For emittance dilution, we calculate  $\Delta(\gamma\epsilon)$ . The projected rms normalized emittance is then

$$\gamma\epsilon = \frac{\gamma\epsilon_0 \sqrt{1 + 2\frac{\Delta(\gamma\epsilon)}{\gamma\epsilon_0}}}{\gamma\epsilon_0 + \Delta(\gamma\epsilon)} \begin{array}{l} \text{before filamentation} \\ \text{after filamentation} \end{array} \quad (2)$$

where  $\gamma\epsilon_0$  is the initial emittance.

Assume the linac consists of thin-lens FODO cells with cell length  $L_{\text{cell}}(s) \propto \gamma(s)^{\alpha}$  and phase advance per cell  $\phi_c$  kept constant. The average  $\beta$ -function over the cell is

$$\bar{\beta} = \frac{L_{\text{cell}}}{2} \left( \cot \frac{\phi_c}{2} + \frac{2}{3} \tan \frac{\phi_c}{2} \right) \quad (3)$$

The wakefunctions of a periodic accelerator structure have been parameterized by a number of authors [4, 5, 6]. Here, we will present simple expressions that are based on the theoretical

short-range dependence plus an empirical term to describe the behavior at longer distances:

$$W_{\parallel}(s) \approx \frac{Z_0 c}{\pi a \sqrt{a^2 + 8.6 s \lambda_{\text{rf}}}} \quad (4)$$

$$W_{\perp}(s) \approx \frac{2 Z_0 c s}{\pi a^2 (a^2 + s \lambda_{\text{rf}})} \quad (5)$$

where  $a$  = iris radius of the structure,  $\lambda_{\text{rf}}$  = rf wavelength. This model holds over the range  $s \lesssim a/4$  and  $0.1 \lesssim a/\lambda_{\text{rf}} \lesssim 0.2$ .

$\phi_{\text{rf}}$  is chosen to minimize the energy spread (see also Sec.2.4.1). The residual energy spread after compensation is found from the convolution of the bunch with the longitudinal wakefield, and the acceleration rf,

$$\frac{\Delta E}{E} \approx \frac{1.25}{G \cos \phi_{\text{rf}}} \left[ 2\pi \epsilon_0 N_B r_e W_{\parallel}(1.5\sigma_z) - \frac{3}{2} \sigma_z \kappa_{\text{rf}} G \sin \phi_{\text{rf}} \right] \quad (\text{FWHM}) \quad (6)$$

where  $G = e \mathcal{E}_{\text{rf}} / (m_e c^2)$ .

There are other techniques of reducing  $\Delta E/E$ . One example is to shape the longitudinal beam distribution [9]; this technique was used in the SLC linac to reduce  $\Delta E/E$  by a factor of  $\sim 2-3$  [10]. Another approach is to rotate the bunch in longitudinal space by  $> 90^\circ$  part way through linac. This flips the sign of the energy correlation and then the longitudinal wakefield can remove the energy spread that it had previously introduced [11]; this technique allows control of  $\Delta E/E$  of very short bunches.

Autophasing [7, 8] can be used to control the single-bunch beam break-up effect (Sec.2.4.1). In the case of a FODO lattice, this requires

$$\sigma_{\delta, \text{auto}} \approx \pi \epsilon_0 N_B r_e W_{\perp}(2\sigma_z) \frac{\bar{\beta} L_{\text{cell}}}{4\gamma \tan(\phi_c/2)} \quad (7)$$

$\sigma_{\delta, \text{auto}}$  scales as  $\gamma^{2\alpha-1}$  and is constant along the linac for  $\alpha = 1/2$ .

Eq.(7) is derived for a rigid offset of the bunch. A similar condition can be found for a bunch offset generated by a misaligned accelerator section. Here, the offset of the bunch depends on  $z$ ; the autophasing condition (condition to preserve  $x-z$  correlation as the bunch propagates down the linac) is  $\approx 1/6$  of Eq.(7) [12].

**Emittance dilution effects, single bunch** [13]  
The primary sources of dilution are transverse wakefields, dispersive errors, and betatron coupling. Additional contributions include rf deflections, beam-gas scattering, quadrupole wakefields, and synchrotron radiation.

**Injection Errors** A beam injected with a trajectory error  $(y_0, y'_0)$  performs a betatron oscillation along the linac. If the autophasing condition is met, the effect of the transverse wakefields will be minimal. But, if the injected beam has an uncorrelated energy spread, it will filament because of the chromatic dependence of the phase advance. The emittance dilution due to this error is

$$\Delta(\gamma\epsilon) \approx \frac{\gamma}{2} \left[ \frac{(1 + \alpha_y^2)}{\beta_y} y_0^2 + 2\alpha_y y_0 y'_0 + \beta_y y'^2_0 \right] \quad (8)$$

where  $\beta_y$  and  $\alpha_y$  are evaluated at the injection point. The case without autophasing energy spread is discussed in Sec.2.4.1.

A similar effect arises if the beam is injected with a  $\beta$ -function mismatch [14],

$$\Delta(\gamma\epsilon) \approx \gamma\epsilon(B_{\text{mag}} - 1) \quad (9)$$

with

$$B_{\text{mag}} \equiv \frac{1}{2} \left[ \left( \frac{\beta}{\beta_{\star}} + \frac{\beta_{\star}}{\beta} \right) + \left( \alpha_{\star} \sqrt{\frac{\beta}{\beta_{\star}}} - \alpha \sqrt{\frac{\beta_{\star}}{\beta}} \right)^2 \right] \quad (10)$$

Subscripts  $\star$  denote mismatched lattice parameters describing the beam. Parameters without  $\star$  denote the natural lattice functions of the linac.

**Accelerator Misalignments** Emittance dilution due to random accelerator structure misalignments is

$$\Delta(\gamma\epsilon) \approx \langle y_a^2 \rangle [\pi \epsilon_0 N_B r_e W_{\perp}(2\sigma_z)]^2 \times \frac{L_{\text{acc}} \bar{\beta}_i}{2\alpha G} \left[ \left( \frac{\gamma_f}{\gamma_i} \right)^{\alpha} - 1 \right] \quad (11)$$

where  $L_{\text{acc}}$  is the length of the accelerator structures (assumed constant along the linac),  $\bar{\beta}_i$  is the initial average  $\beta$ -function.

Equation (11) assumes that the alignment of the individual structures is random. In case all the structures between a pair of quadrupoles have a systematic misalignment (e.g. when quadrupoles are used for alignment reference),

$$\Delta(\gamma\epsilon) \approx \langle y_a^2 \rangle [\pi \epsilon_0 N_B r_e W_{\perp}(2\sigma_z)]^2 \times \frac{L_{\text{cell } i} \bar{\beta}_i}{4\alpha G} \left[ \left( \frac{\gamma_f}{\gamma_i} \right)^{2\alpha} - 1 \right] \quad (12)$$

where  $L_{\text{cell } i}$  is the initial cell length. In general, Eq.(12) contributes more than Eq.(11) because more structures are contributing.

**BPM Misalignments** Consider a 1-to-1 trajectory correction where one minimizes the readings on BPMs located at all of the quadrupoles. If these

BPM's are misaligned, the trajectory will be offset in the accelerator structures and quadrupoles. Assuming that the autophasing condition is met, one has

$$\Delta(\gamma\epsilon) \approx \langle y_{\text{BPM}}^2 \rangle [\pi\epsilon_0 N_B r_e W_{\perp}(2\sigma_z)]^2 \times \frac{\cos \frac{\psi_c}{2}}{\sin^3 \frac{\psi_c}{2}} \frac{L_{\text{cell}}^2}{16\alpha G} \left[ \left( \frac{\gamma_f}{\gamma_i} \right)^{2\alpha} - 1 \right] \quad (13)$$

If the BPM readings are only located at the focusing quads, and the defocusing quads are well aligned,

$$\Delta(\gamma\epsilon) \approx \langle y_{\text{BPM}}^2 \rangle [\pi\epsilon_0 N_B r_e W_{\perp}(2\sigma_z)]^2 \tan^3 \frac{\phi_c}{2} \times \left( 1 - \sin \frac{\phi_c}{2} \right) \frac{L_{\text{cell}}^2}{36\alpha G} \left[ \left( \frac{\gamma_f}{\gamma_i} \right)^{2\alpha} - 1 \right] \quad (14)$$

Eq.(14) contributes significantly less than Eq.(13) because the trajectory offset is demagnified in the defocusing quads, leading to smaller offsets on average.

Effect of misalignments of the defocusing quads can be included by multiplying Eq.(14) by

$$1 + 4 \left( \frac{y_{\text{QD}}}{y_{\text{BPM}}} \right)^2 \left( 1 - \sin \frac{\phi_c}{2} \right)^2 \tan^2 \frac{\phi_c}{2}$$

where  $y_{\text{QD}}$  is the misalignment of the defocusing quads.

**Betatron Coupling** In future linear colliders, typically  $\epsilon_y \ll \epsilon_x$ , and there are tight tolerances on the allowable skew quad fields. Assuming purely random errors [3],

$$\Delta(\gamma\epsilon)_y \approx \gamma\epsilon_x 4N_{\text{cell}} \left( \frac{\tilde{K}}{K} \right)^2 \quad (15)$$

where  $\tilde{K}$  is the skew quad gradient normalized by the magnetic rigidity  $B\rho$  and  $K$  is the normalized quad gradient. For roll errors,  $\tilde{K} = 2\Theta_q K$ .  $N_{\text{cell}}$  is the total number of FODO cells.

Similarly, assuming systematic errors,

$$\Delta(\gamma\epsilon)_y \approx \gamma\epsilon_x \left( \frac{\tilde{K}_{QF} + \tilde{K}_{QD}}{|K_1|} \right)^2 \frac{\sin^2 N_{\text{cell}} \frac{\Delta\phi_c}{2}}{\sin^2 \frac{\Delta\phi_c}{2}} \quad (16)$$

where subscripts  $QD$  and  $QF$  denote the skew components of the focusing and defocusing quads,  $\Delta\phi_c = \phi_{xc} - \phi_{yc}$  is the difference between the  $x$ - and  $y$ -phase advances per cell. Note that  $\Delta(\gamma\epsilon)_y \propto N_{\text{cell}}^2$  when  $\Delta\phi_c \lesssim 2/N_{\text{cell}}$ . It is thus useful to split  $\phi_{xc}$  and  $\phi_{yc}$  by a few degrees. Also note that the dilution depends on the sum  $\tilde{K}_{QF} + \tilde{K}_{QD}$ . If all elements have the same error,

i.e., the same roll or same pole error, the dilution will be small because the errors cancel.

**RF Deflections** Rf deflections occur if the time varying acceleration field is not oriented in the direction of beam propagation. The misalignment can arise from a misaligned accelerator structure, an angular trajectory through a structure, or asymmetries in a structure such as tilted irises or the couplers. In addition to deflecting the beam, the rf deflections will increase the projected emittance by causing a deflection that is a function of  $z$ ,

$$\Delta(\gamma\epsilon) \approx \langle g^2 \rangle (\sigma_z k_{\text{rf}})^2 \frac{\bar{\beta}_i L_{\text{acc}} G}{\alpha} \left[ \left( \frac{\gamma_f}{\gamma_i} \right)^\alpha - 1 \right] \quad (17)$$

where  $g$  is the longitudinal-transverse coupling of the deflection,

$$g = \begin{cases} \Theta_{\text{acc}}/2 & \text{structure misalignment} \\ \Theta_{\text{tra}}/2 & \text{trajectory misalignment} \\ \lesssim \Theta_{\text{iris}} & \text{tilted irises} \end{cases} \quad (18)$$

**Beam-Gas Scattering** Beam-gas scattering is rarely an issue for the emittance in a linac,

$$\Delta(\gamma\epsilon)_y \approx n_{\text{gas}} \frac{160\pi Z(Z+1)r_e^2 \bar{\beta}_i}{\alpha G} \left[ \left( \frac{\gamma_f}{\gamma_i} \right)^\alpha - 1 \right] \quad (19)$$

where  $Z$  is the atomic number of the gas and  $n_{\text{gas}}$  is the density of the residual gas. At 20°C,  $n_{\text{gas}} \approx 3.22 \times 10^{22} n_{\text{mol}} P [\text{m}^{-3}]$ . Here,  $n_{\text{mol}}$  is the number of atoms per molecule of gas and  $P$  is the vacuum pressure in torr.

The primary effect is to contribute to a halo. The number of particles scattered out to an amplitude greater than  $k$  times the rms beam size can be estimated,

$$\frac{\Delta N}{N} \approx \frac{n_{\text{gas}}}{k^2} \frac{\bar{\beta}_i}{\gamma\epsilon} \frac{4\pi Z^2 r_e^2}{\alpha G} \left[ \left( \frac{\gamma_f}{\gamma_i} \right)^\alpha - 1 \right] \quad (20)$$

### Emittance dilutions effects, multi-bunch

**Beam Break-up** Multi-bunch BBU [15, 16, 17] leads to an amplification of the incoming trajectory jitter. One solution is reduction of the long-range transverse wakefield in the structure design. Assuming the daisy chain model (Sec.2.4.1), the criterion for little or no blowup for the case  $\alpha = 1/2$  is [18]

$$\left| \frac{2\pi\epsilon_0 N_B r_e W_{\perp}(S_B) \bar{\beta}_i}{\alpha G} \right| \left( \frac{\gamma_f}{\gamma_i} \right)^\alpha < 1 \quad (21)$$

where  $W_{\perp}(S_B)$  is the wakefield at the following bunch.

Assuming the linac is operated with minimized multi-bunch BBU and close to the autophasing condition, one can estimate the emittance dilution due to misaligned accelerator structures and a corrected trajectory similarly to the single bunch case. One can use the single bunch emittance dilution formulas, simply replacing  $W_{\perp}(2\sigma_z)$  with  $4W_{\text{sum}}$ , where  $W_{\text{sum}}^2 \equiv$

$$\frac{1}{n_B} \sum_{i=0}^{n_B-1} W_{\perp}^2(iS_B) - \left[ \frac{1}{n_B} \sum_{i=0}^{n_B-1} W_{\perp}(iS_B) \right]^2 \quad (22)$$

$n_B$  is the number of bunches.

**Component jitter and stability** There are three issues: (i) jitter where the beam trajectory changes from pulse-to-pulse, (ii) emittance degradation due to shifts in the position of the accelerator components, and (iii) the degradation of the diagnostic resolution that arises from the beam centroid jitter which can lead to poor convergence of tuning or correction procedures. We consider issues (i) and (ii). Issue (iii) depends upon the diagnostics and the operational procedures.

Sources of vibration include natural seismic motion and man-made cultural noise (Sec.5.13). The ground motion and vibration can be divided into three regimes: high frequency where there is little or no spatial correlation of the vibration, the intermediate regime where the lower frequency ground motion tends to be relatively well correlated, and slow drifts where the motion is uncorrelated.

In general, the effect of the ground vibration is described with a 2-D power spectrum  $P(\omega, k)$  [19]. The motion of the beam at a point can be described by

$$\langle \Delta y(t)^2 \rangle = \frac{1}{(2\pi)^2} \int P(\omega, k) G(k) F(\omega) d\omega dk \quad (23)$$

where  $G(k)$  is the spatial response function for the focusing lattice, and  $F(\omega)$  is the temporal response of the feedback systems. The growth of the beam spot size can be evaluated in a similar manner with a different function  $G$ .

Evaluating Eq.(23) requires detailed computer calculation, but one can estimate the two limiting extremes: uncorrelated high-frequency jitter and slow uncorrelated drifts of components which can be described with the “ATL” relation, which states that after a time  $T$ , the relative motion of two points, separated by a distance  $L$ , is

equal to [20]

$$\langle \Delta y^2 \rangle = A \times T \times L$$

The high-frequency jitter can be used to estimate the motion of the beam centroid while the ATL-drifts primarily result in an increase in the beam emittance, since the actual motion of the centroid can be compensated with beam-based feedback systems, and thereby specify the time interval between component re-alignment.

Uncorrelated vibration of the linac quads  $y_q$  will deflect the trajectory, causing a trajectory motion, at the end of the linac, of

$$\langle \Delta y^2 \rangle \approx \langle y_q^2 \rangle \frac{4\gamma_i N_{\text{cell}} \tan \frac{\phi_c}{2}}{\gamma \epsilon_y L_{\text{cell } i}} \left[ \left( \frac{\gamma_f}{\gamma_i} \right)^{1-\alpha} + 1 \right] \quad (24)$$

Similarly, if uncorrected, the ATL motion will cause a trajectory displacement of

$$\langle \Delta y^2 \rangle \approx 8AT \frac{N_{\text{cell}} \gamma_f}{\gamma \epsilon_y} \frac{\sin^2 \frac{\phi_c}{4}}{\sin \phi_c} \frac{1-\alpha}{2-\alpha} \quad (25)$$

where  $A$  depends only the specifics of the site but typically  $A$  ranges from 0.1 to 100 nm<sup>2</sup>/m.s. Of course, this trajectory drift could be corrected with beam-based feedback systems at the end of the linac (or portion of the linac with multiple feedback systems) but it will still result in emittance degradation. The emittance increase due to the trajectory oscillation is

$$\langle \Delta(\gamma\epsilon) \rangle \approx 8AT N_{\text{cell}}^3 \gamma_i \delta_e^2 \frac{\sin^2 \frac{\phi_c}{4}}{\cos^2 \frac{\phi_c}{2} \sin \phi_c} \frac{1-\alpha}{2-\alpha} \quad (26)$$

where  $\delta_e$  is the rms deviation from the autophasing energy spread. Typically,  $\delta_e \sim 20 \sim 30\%$  of  $\delta_{\text{auto}}$  since the energy variation along the bunch, induced by the rf and longitudinal wakefield, usually does not exactly match the desired form unless the bunch charge distribution is specially shaped.

**Distributions** The preceding estimates have been for the *expected* emittance dilution due to a random distribution of errors. For a specific set of errors, the dilution can deviate from the expected value significantly.

Assuming a Gaussian distribution of errors which cause a *single* dominant cause of dilution, the dilution will have the distribution [3],

$$f(\Delta(\gamma\epsilon)) = (1/\mu) e^{-\frac{\Delta(\gamma\epsilon)}{\mu}} \quad (27)$$

Here,  $\mu$  is the expected dilution. Strictly, Eq.(27) applies only when the distribution of errors is

Gaussian, but in a long linac with many errors, is applies approximately regardless of the detailed distribution for the errors.

Distribution (27) has a long tail. There is a 5% probability that any specific case will exceed  $3\mu$ . These large dilutions arise when the errors add in phase at the betatron frequency. Fortunately, this situation is easy to measure and correct by advanced correction techniques (see later).

If there are  $N$  forms of emittance dilution that are all contributing to the total dilution with equal magnitude, the distribution function for the emittance is described with a  $\chi^2$ -squared distribution with  $2N$  degrees-of-freedom. For example, a total emittance dilution of 50% would have a  $\chi^2$ -squared distribution with 4-d.o.f. if it consisted of 25% dilution due to transverse wakefields and 25% dilution due to betatron coupling. As  $N$  increases, the probability of having an emittance much larger than the expected value decreases. It is difficult to calculate the distribution function in the general case, but typically one obtains a close approximation with  $N = 1$  or 2.

**Advanced trajectory and emittance correction techniques** A number of more advanced beam-based techniques have been developed or proposed to reduce emittance dilution. The most straightforward is to use “ $\epsilon$ -bumps”. Here, the measured beam emittance is minimized by adjusting some dilution source upstream of the measurement point. For example, one can reduce the dilution due to the transverse wakefields by appropriately misaligning a few accelerator structures. Such techniques are “non-local” because the sources of dilution are not reduced; instead, additional dilutions are added to cancel the effect of the others.

The  $\epsilon$ -bump technique is routinely used at the SLC to reduce  $\Delta(\gamma\epsilon)$  by  $\sim$  an order of magnitude [21]. Some of the difficulties with the non-local correction techniques are: (i) the beam emittance and tails need to be measured accurately, (ii) multiple measurement stations may be needed to prevent the dilutions from filamenting, and (iii) being nonlocal, it is sensitive to small changes in betatron phase advance that can arise from changes in the quadrupole strengths or energy profile; this results in large fluctuations in the beam emittance at the SLC over a time scale of hours.

“Local” correction techniques measure and correct the emittance dilution sources locally. Typically, this is done by measuring the difference

of two, or more, trajectories while some parameter is varied. For example, to align the quads and minimize the residual dispersion, one can measure the trajectory while changing the quad power supplies or the beam energy. A few of these steering techniques were tested at the SLC [22, 23], including the “Dispersion-Free” [24] and the “Wake-Free” [25] algorithms. A version of the Dispersion-Free steering technique was implemented as part of the standard trajectory correction on the SLC. Similarly, to measure the short-range transverse wakefield due to accelerator structure misalignments, one can vary the bunch intensity or length [12, 26], and, to measure the long-range transverse wakefield due to accelerator structure misalignments, one can vary the bunch charge or the bunch spacing [27].

In all cases, the best alignment that can be attained using local techniques is determined by the BPM or diagnostic precision and is roughly independent of the magnitude of the misalignments; this arises because the techniques measure the differences between trajectories. To estimate the effectiveness of these techniques, there are analytic approaches [28, 29], or one can perform computer simulations which include many of the envisioned errors. This is important because the techniques frequently rely on measuring small differences between the beam trajectories and can be sensitive to beam jitter or power supply fluctuations. Finally, another difficulty with these approaches is that the beam emittance is not actually measured and, instead, the dilution sources are inferred from other measurements—thus, it is desirable to include emittance diagnostic stations and use non-local correction to remove any residual dilutions.

## References

- [1] R.D. Ruth, Proc. 1986 CERN/US Part. Acc. School (1986)
- [2] H. Henke, LINAC 88
- [3] T.O. Raubenheimer, PhD Thesis, Stanford U. (1991)
- [4] P. Wilson, AIP Proc. 87 (1982) p.450
- [5] R.B. Palmer, Ann. Rev. Nucl. Part. Sci. 40 (1990) 529
- [6] K. Yokoya, private communication (1995)
- [7] V. Balakin, SLAC-355 (1989) p.55
- [8] V. Balakin et al, HEACC 83, 119
- [9] G.A. Loew, J.M. Wang, PAC 85, 3228
- [10] F.J. Decker et al, LINAC 94, 47
- [11] K.L. Bane et al, PAC 93, 3933

### Sec.4.3: LINAC OPERATION

- [12] T.O. Raubenheimer, K. Kubo, NIM A370 (1996) 303
- [13] T.O. Raubenheimer, PAC 93, 11
- [14] F.J. Decker et al, PAC 91, 6
- [15] R. Helm, G.A. Loew, Linear Accelerators, N. Holland (1970)
- [16] R.L. Gluckstern, R.K. Cooper, P.J. Channell, PA 16 (1985) 125
- [17] K. Yokoya, DESY 86-084 (1986)
- [18] K. Thompson, R. Ruth, PR D41 (1990) 964
- [19] For example see: A. Sery, O. Napoly, PR E53 (1996) 5323
- [20] V. Shiltsev, 4th Int. Workshop on Acc. Align. (1995)
- [21] J.T. Seeman, F.J. Decker, I. Hsu, Proc. 15th Int. Conf. High Energy Acc. (1992) 879
- [22] C. Adolphsen et al, PAC 89, 977
- [23] R. Assmann et al, KEK Proc. 95-12 (1995)
- [24] T.O. Raubenheimer, R.D. Ruth, NIM A302 (1991) 191
- [25] T.O. Raubenheimer, NIM A306 (1991) 61
- [26] C. Fischer, PAC 95, 731
- [27] R. Assmann et al, PAC 97, 503
- [28] C. Adolphsen, T.O. Raubenheimer, PAC 93, 417
- [29] A. Sery, A. Mosnier, DAPNIA/SEA-96-06 (1996)

#### 4.3.2 Operation of Superconducting Linacs

*S. Schreiber, DESY*

The operation of Superconducting (SC) linear accelerators (linacs) is illustrated by the example of FLASH.

**Superconducting accelerator FLASH**  
 FLASH is the free-electron laser at DESY in Germany in user operation since 2005. It produces laser-like radiation in the VUV and soft x-ray wavelength range with unprecedented brilliance.[1, 2] The radiation pulses in the wavelength range of 4.1 to 47 nm are extremely short, from 10 to 100 fs, with a peak power exceeding 1 GW. The FEL is driven by a high quality electron beam with an energy of up to 1.25 GeV. FLASH uses TESLA superconducting acceleration technology.

State-of-the-art superconducting cavities have accelerating gradients well above 25 MV/m. At the same time, the (unloaded) quality factor  $Q_0$  is very high exceeding  $10^{10}$ . A high quality factor allows efficient operation with high duty cycle.

The quality factor is proportional to the ratio of stored energy  $W$  to the energy lost in one rf

cycle:

$$Q_0 = 2\pi \frac{f_0 W}{P_{\text{diss}}} = \frac{G}{R_{\text{surf}}}, \quad (1)$$

$f_0$  is the resonance frequency,  $P_{\text{diss}}$  the dissipated rf power.  $Q_0$  can also be expressed as the ratio of a geometrical factor  $G$  (with a dimension of resistance) to the surface resistance  $R_{\text{surf}}$ .

The TESLA type 9-cell cavities used at FLASH are built with solid niobium and have a length of 1.038 m. They are operated at 1.3 GHz in the fundamental mode TEM<sub>010</sub> ( $\pi$ -mode). Higher harmonics are sufficiently damped using higher-order mode couplers. With a geometrical factor of  $G = 270 \Omega$  and a residual surface resistance of less than  $10 \text{n}\Omega$ , the quality factor is well above  $2 \times 10^{10}$ . The design  $R/Q$  is  $518 \Omega$  (using the definition of dissipated power  $P_{\text{diss}} = V_{\text{cav}}^2 / R_{\text{sh}}$  with cavity voltage  $V_{\text{cav}}$  and shunt impedance  $R_{\text{sh}}$ ). For a comprehensive overview of TESLA type superconducting cavities see [3].

The loaded quality factor  $Q_L$  is given by

$$Q_L = \frac{f_0}{\Delta f_L} \quad (2)$$

with the (loaded) bandwidth  $\Delta f_L$ . The bandwidth is defined as the full width at half height of the resonance curve. At FLASH, the coupling of the rf to the cavity is adjusted such that  $Q_L = 3 \times 10^6$  with a loaded bandwidth of  $\Delta f_L = 430 \text{ Hz}$ . For superconducting cavities, the loaded  $Q_L$  is much smaller than the unloaded  $Q_0$ , so that the coupling  $\beta$ , defined as  $\beta = Q_0/Q_L - 1$ , is with  $\approx 3000$  much larger than 1 – in contrast to normal conducting cavities. Therefore, the properties of superconducting cavities are mainly determined by the coupling or the loaded  $Q_L$ , while the cryogenic load is determined by the unloaded  $Q_0$ .

**Accelerating module** Eight cavities are assembled in a 12.2 m long cryostat module cooled with superfluid helium. Operation at 2 K is assured by stabilization of the helium pressure at 31 mbar to the  $3 \times 10^{-4}$  level (rms). The cryostat design takes care that the static heat load, important for efficient operation, is less than 1 W/m.

Superconducting technology allows acceleration of many bunches grouped in bunch trains with a large number of bunches per second. FLASH is designed to accelerate a beam current of up to 9 mA – within a bunch train of 800  $\mu\text{s}$  in length, at 10 Hz repetition rate. For a single bunch charge of 1 nC, this leads to 72000 accelerated bunches per second.

After the upgrade in 2009/10, FLASH operates seven modules. Two modules with together 16 cavities are powered by one klystron. A 5 MW klystron is sufficient to provide a drive rf power of 250 kW per cavity required for 25 MV/m operational gradient. The first module has its own 5 MW klystron, the last two modules exceeding 25 MV/m are powered with a 10 MW multi-beam klystron. The energy reach of FLASH is 1.25 GeV. The electron source is a photoinjector with a warm L-band copper rf gun together with a suitable drive laser providing low emittance beams.

Important for successful acceleration of bunch trains with hundreds of bunches is the stability and flatness of the accelerating gradient and phase along the rf pulse, together with a reliable exception handling and machine protection system to prevent catastrophic beam losses.

**Stability of acceleration** There are several sources of errors which affect the amplitude and phase of the accelerating field – such as errors in cavity tuning, Lorentz force detuning, microphonics, errors and fluctuations of the drive amplitude and phase, errors of relative phase of beam and cavity rf for individual cavities, errors in injection phase. Also the stability of the total beam charge affects the stability of the beam energy, phase, and arrival time.

Noise from the environment induces vibrations of the cavities (microphonics). The frequencies of microphonics are typically below 1 kHz, which does not much affect the acceleration within a bunch train, but may lead to fluctuations of the beam energy and phase from train to train.

A feedback system stabilizing amplitude and phase within a pulse train and from train to train is mandatory. The so called low-level rf system analyses the signals from pick-ups installed in each cavity to measure amplitude and phase of the accelerating field. Since one klystron drives many cavities (16 at FLASH), the system calculates the vector sum of amplitude and phase, and provides feedforward and feedback for stabilization. The goal is to achieve stable acceleration of all bunches along the pulse train and from train to train at the  $10^{-4}$  level. Low latency and fast real time computation abilities with high performance ADC's are required. FLASH uses modern FPGA based controllers with 14 bit ADC's and DAC's with 81 MHz clock rate.

Many noise signals distorting the amplitude and phase flatness are repetitive from rf pulse to rf pulse. For instance drifts along the pulse due to thermal loads, beam loading effects, Lorentz-Force detuning, and others. A learning feed-forward algorithm recognizes the repetitive distortions and compensates them applying appropriate feedforward signals. The energy stability achieved at FLASH during operation with long bunch trains is better than  $10^{-4}$  (rms) from train to train and along the bunch train. The phase stability is better than  $0.02^\circ$ .

Stability in arrival time is important for free-electron lasers, since pump and probe experiments rely on a measurable delay between the pump and probe beam to the femtosecond level. The goal is to achieve an arrival time stability shorter than the radiation pulse length or, in other words, at the 10 fs level.

An already excellent energy stability of  $10^{-4}$  of the beam entering the first bunch compressor at FLASH with an  $R_{56}$  of 0.18 m induces an arrival time jitter of 60 fs. The arrival time change  $\delta t$  of a bunch with a momentum deviation of  $\delta p$  from the nominal momentum  $p$  is given by  $\delta t = R_{56}/c \cdot \delta p/p$ . Other sources like the drive laser and rf gun add to the arrival time jitter. The measured arrival time jitter at FLASH is usually 80 fs (rms) after the first bunch compressor.

A reduction of the arrival time jitter to the 10 fs level is difficult to achieve with the low-level rf feedback alone, and a beam based feedback is required. The beam based feedback system at FLASH measures the arrival time of the bunches by comparing the signals obtained by a pick-up antenna with an ultra-stable fiber laser. The amplitude of acceleration upstream of the bunch compressor is corrected for subsequent bunches in the train. The resolution achieved is 6 fs [5]. So far, promising results have been achieved, and the arrival time has been stabilized to within 20–40 fs.

**Beam loading** The forward rf power  $P_{\text{fwd}}$  required to accelerate a beam with current  $I_b$  at phase  $\phi$  and detuning  $\Delta f$  is [4]

$$P_{\text{fwd}} = \frac{V_{\text{cav}}^2}{8R_L} \left[ \left( 1 + \frac{2R_L I_b}{V_{\text{cav}}} \cos \phi \right)^2 + \left( 2Q_L \frac{\Delta f}{f_0} + \frac{2R_L I_b}{V_{\text{cav}}} \sin \phi \right)^2 \right] \quad (3)$$

with the loaded resistance  $R_L = (R/Q)Q_L$ . For the case of zero beam current, an accelerating gradient of 25 MV/m ( $V_{\text{cav}} = 26$  MV) requires a forward power of 54 kW per cavity, a current of 9 mA requires 288 kW for  $Q_L = 3 \times 10^6$ , on-crest acceleration and no detuning.

The coupling of the rf to the cavities is adjusted to optimize the drive power for a given beam current. The couplers at FLASH are usually optimized for 9 mA operation. The optimum coupling is such that the sum of beam induced voltage and cavity voltage is matched to the generator for minimum rf power and zero reflected power (see Sec.2.4.3.2). A derivation of the optimum coupling is given in [4] and can be expressed by

$$Q_L^{\text{opt}} = \frac{V_{\text{cav}}}{2\frac{R}{Q}I_b \cos(\phi)}. \quad (4)$$

For a cavity voltage of  $V_{\text{cav}} = 26$  MV (25 MV/m) and on-crest acceleration ( $\phi = 0$ ) the optimum coupling for a beam current of 9 mA is  $Q_L^{\text{opt}} = 2.8 \times 10^6$ . Optimizing the loaded quality factor reduces the forward power needs from 288 kW to 233 kW.

Filling a cavity on resonance with constant forward power yields in an increase in cavity voltage  $V_{\text{cav}}$  by [4]

$$V_{\text{cav}}(t) = 2R_L I_g \left(1 - e^{-t/\tau}\right) \quad (5)$$

with a filling time of  $\tau = 2Q_L/\omega_0$ . The generator current  $I_g$  is given by

$$I_g = 2 \sqrt{\frac{P_g}{2\frac{R}{Q}Q_L}}. \quad (6)$$

The beam with a current  $I_b$  induces a voltage in the cavity of

$$V_b(t) = -2R_L I_b \left(1 - e^{-(t-t_{\text{inj}})/\tau}\right) \quad (7)$$

The optimum injection time to achieve immediate steady state operation is  $t_{\text{inj}} = \tau \ln 2$ , then when the cavity voltage reaches half of its maximum. The cavity field at the optimum injection time is

$$V_{\text{cav}}(t = t_{\text{inj}}) = 2\frac{R}{Q}Q_L I_b. \quad (8)$$

To summarize, for optimized operation with a beam current of 9 mA and an accelerating gradient of 25 MV/m, we need a forward power of

$P_{\text{fwd}} = 233$  kW,  $Q_L = 2.8 \times 10^6$  and get a generator current of  $I_g = 18$  mA, a filling time of  $\tau = 681 \mu\text{s}$ , and an optimal time to inject the beam of 472  $\mu\text{s}$  after start of rf power. The single bunch transient, the voltage drop for a bunch with a charge of 1 nC is  $1.6 \times 10^{-4}$ .

In practice, FLASH is operated with many different bunch charges and bunch train currents, ranging from 100 pC to 2 nC, and currents from 10  $\mu\text{A}$  to 4.5 mA. The rf power needs to be adjusted according to the actual current. This is done by measuring the bunch train current with toroids (resolution 3 pC) and applying an appropriate feedforward rf power (beam loading compensation).

**Lorentz force detuning** The Lorentz force between the rf magnetic field and the induced currents in a thin surface layer of the cavity walls leads to a mechanical deformation. This induces a detuning proportional to the square of the cavity field  $E_{\text{acc}}$

$$\Delta f = -K_L E_{\text{acc}}^2 \quad (9)$$

with the detuning constant  $K_L = 1 \text{ Hz}/(\text{MV/m})^2$  (for the mechanically stiffened TESLA cavities).

In pulsed operation, the detuning changes dynamically along the rf pulse. Measurements give a linear detuning (rf flat-top) of  $0.5 \text{ Hz}/(\text{MV/m})^2/\text{ms}$  [6]. As a result, in pulsed operation with a field of 25 MV/m, the cavity is detuned by 260 Hz within the pulse train duration of 800  $\mu\text{s}$ . Since this is in the order of the cavity bandwidth, the effect has to be accounted for. Using Eq.(3) with  $\Delta f = 130$  Hz, this would lead to an energy reduction along a 9 mA bunch train of 6.8%. To keep the energy stable, an additional rf power of 7.7 % would be required to compensate the detuning.

For low gradients, pre-detuning together with the low-level rf feedback is sufficient to regulate against Lorentz force detuning. However, for operation with higher gradients above 25 MV/m, the required additional power would be too high; a fast tuning system is required [7]. At FLASH, Piezo-stacks are mounted in parallel to the motorized tuning system of the cavities. They allow a fast tuning within the rf pulse and compensate efficiently the Lorentz force effect.

**Beam loss detection** At FLASH the total power of the design beam is 72 kW, the beam for ILC will exceed 10 MW. A permanent loss of beam of the order of 0.1% into the cryostat would lead

to an unacceptable heat load. For many beam-line components like the SASE undulators, allowed losses are one or two orders of magnitude smaller. Only a few Gy per day are tolerated for the FLASH undulators translating in a tolerable loss of approximately  $5 \times 10^{-5}$  (nominal beam, 10 years undulator lifetime).

Beam losses are detected with two systems [8]. Toroids along the machine measure the charge of the electron bunches with a single bunch resolution of 3 pC. An alarm is triggered, whenever a single bunch loss, or a loss of the integrated charge along a bunch train occurs. A second system is based on beam loss detection using scintillators placed at critical positions along the accelerator. Special care has been taken to protect the undulator system from beam losses larger than  $10^{-5}$ . Sudden losses trigger an alarm blocking the drive laser of the electron source within 3  $\mu$ s. Small permanent losses lead to warning and inhibit beam for a certain period.

In case of sudden beam loss, the beam loading compensation of the low-level rf system is triggered to turn off immediately (exception handling). Otherwise the overhead on rf power would eventually quench the superconducting cavities. A proper exception handling limiting the forward power for individual cavities allows operation close to the quench limit. Since 16 cavities are driven by one klystron and only the vector sum of all cavities is regulated, some cavities may have increasing, others, for compensation, decreasing amplitudes along the rf pulse. Proper tuning to resonance and adjustment of the loaded quality factor is required to keep all amplitudes flat. At FLASH, an operable gradient of about 5 to 10 % below the actual quench limit has been achieved.

## References

- [1] FLASH web site <http://flash.desy.de/>
- [2] W. Ackermann et al., Nature Photonics 1 (2007) 336
- [3] B. Aune et al, PRST-AB 3, 092001 (2000)
- [4] T. Schilcher, PhD Thesis, U. Hamburg, TESLA-1998-20 (1998)
- [5] F. Löhl et al, PRL 104, 144801 (2010)
- [6] M. Liepe, PhD thesis, U. Hamburg, DESY-THESIS-2001-045 (2001)
- [7] M. Liepe, W.D. Moeller, S.N. Simrock, PAC 01, 1074
- [8] L. Froehlich, PhD thesis, U. Hamburg, DESY-THESIS-2009-012 (2009)

### 4.3.3 Halo in High Intensity Proton Linacs

*T.O. Raubenheimer, SLAC*

In high intensity proton linacs, beam loss must be carefully controlled to avoid activation of the accelerator components. This requires controlling the beam halos. Halos are most commonly generated during bunching when rapidly changing space charge forces can generate transverse tails and incomplete bunching can generate longitudinal tails. In addition, halos can be populated when the space charge force is modulated due to phase space mismatches which cause beam density oscillations [1, 2]; this modulation of the space charge force can resonantly drive particles out to large transverse amplitudes. Relatively simple models [3, 4, 5, 6] predict the qualitative behavior and show good agreement with detailed simulation models.

## References

- [1] M. Reiser, PAC 91, 2497
- [2] A. Cucchetti et al, PAC 91, 251
- [3] J. O'Connell, PAC 93, 3657
- [4] J. Lagniel, NIM A345 (1994) 46
- [5] R. Gluckstern, PRL 73 (1994) 1247
- [6] T. Wangler et al, LINAC 96, 372

## 4.4 RECIRCULATED ENERGY RECOVERY LINACS

*S.A. Bogacz, D.R. Douglas, G.A. Krafft,  
Jefferson Laboratory*

### 4.4.1 Recirculation and Energy Recovery

Linacs that are recirculated have several advantages in achieving electron beam parameters outside of the scope of the traditional ring accelerators or linacs [1]. Beam quality emerging from a recirculated linac is not limited by synchrotron radiation effects as in electron storage rings; with the advent of energy recovery, currents can be raised to within an order of magnitude of the best lepton storage rings in existence now.

The original impetus for recirculated superconducting linac development came from the electron scattering community of nuclear physicists [2]. Coincidence measurements involving multiple-particle final states are most advantageously run at limited peak currents to reduce effects of accidental coincidences. To maintain

high event rates, integrated at low peak current, high average current is desired, implying c.w. operation. Early c.w. superconducting linacs were built to perform electron scattering experiments at University of Illinois and Stanford University. These devices were upgraded in energy by the simple expedient of multipass beam recirculation where beam emerging from the end of the linac is taken back to the beginning of the linac and run at the same accelerating phase as the first pass [3, 4]. More recently, a recirculated superconducting linac was built at Darmstadt, initially for electron scattering experiments [5]. Multipass beam recirculation has also been achieved at normal conducting linacs at Massachusetts Institute of Technology's Bates Laboratory [6], the United States National Bureau of Standards [7], the Mainz cascaded racetrack microtrons [8] and at the Budker Institute [9]. Similarly, economic considerations drove the design of the 6 GeV CEBAF c.w. electron accelerator at Jefferson Lab [10] to multipass recirculation, where a relatively expensive superconducting linac could be replaced by relatively inexpensive recirculation beam lines to reduce the overall project cost.

Aside from economic considerations, applications requiring the following elements should generally be well suited to deploying a recirculated and/or energy recovered linac: c.w. or other high duty factor operation, high beam average current, low delivered beam energy spread, and low delivered beam emittance. C.w. beam acceleration with high accelerating gradient (5-10 MV/m) generally requires deploying a (recirculated) linac consisting of superconducting accelerator structures. GeV-scale recirculated linacs at 100 mA average current would ordinarily require at least 100 MW of installed rf power merely to accelerate the beam load. Beam energy recovery allows the rf beam loading of the cavities to be substantially lowered. In applying this idea with a back to front beam recirculation, the beam recirculation path length is chosen to be a half integral number of rf wavelengths long. Because the beam sees an accelerating phase on the lower accelerating beam passes through the linac, after a phase shift of 180 degrees energy it is delivered back to the (SC) rf cavity on the higher decelerating beam passes, energy is extracted and transferred directly to the accelerating beams without the need for power to be provided by the rf sources [11]. To the extent that the average beam load from the accelerating passes completely cancels

the beam load from the decelerating beam passes, there is no limit to the average current that may be accelerated due to rf source capacity. An alternate method to provide energy recovery is to recirculate the beam back through the accelerating cavities in the opposite direction [12]. Because electrons are relativistic and travel at the velocity of light largely independently of beam energy, coordinating beam acceleration and energy recovery over a linac consisting of multiple accelerating cavities is straightforward for electrons, but difficult for non-relativistic particles. Because the beam transit time through the recirculated linac is much smaller than the radiation-induced emittance growth times in the bending arcs, the beam longitudinal and transverse emittances can be much smaller in energy recovered linacs than in storage ring accelerators that operate at the same energy. It should also be noted that energy recovery is an important element in the design of high average current electrostatic accelerators.

Beam energy recovery was first proposed as a way to construct high luminosity colliders for high energy physics [13]. Although never realized in this application, energy recovered accelerators have been built as electron cooling drivers and high power free electron laser drivers [14, 15, 16]. Many proposed applications utilize the advantages of energy recovered linacs. For example, Cornell University is investigating the energy recovered linac as an undulator driver yielding superior, high average brilliance x-ray sources as an upgrade to their conventional synchrotron light facility [17]. Similar programs exist at Argonne and Daresbury Laboratories [18], and in Japan [19, 20]. Brookhaven National Laboratory and CERN are investigating the use of high average current energy recovery linacs as electron sources for high-luminosity electron ion colliders [21].

**Applications of ERLs** Applications of ERL include high average power FELs (Fig.1), synchrotron light sources, Terahertz and Compton sources as well as electron cooling devices and electron-ion colliders for nuclear and high energy physics. Past and present ERLs are listed in Tab.1. FEL ERLs promise high overall system efficiency, unprecedented average laser power ( $\sim 100$  kW) and reduced beam dump activation. Typical beam parameters for high power FEL ERLs are bunch charge  $\sim 0.1\text{-}1$  nC, average current  $\sim 100$  mA, and energy  $\sim 100\text{-}200$  MeV.

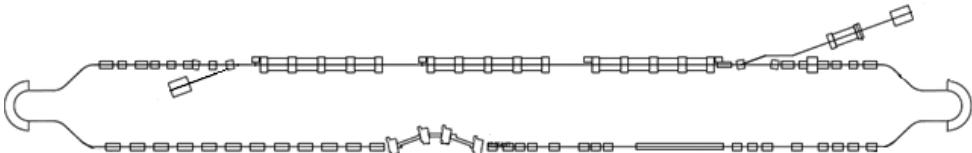


Figure 1: JLab IR-UV Demo - Example of ERL-based FEL.

Table 1: Parameters of operating ERLs.

|                      | E [MeV] | $I_{av}$ [mA] | $Q_b$ [pC] | $\epsilon_N$ [ $\mu m$ ] | Rep [MHz] | Duty [%] |
|----------------------|---------|---------------|------------|--------------------------|-----------|----------|
| Chalk River [40]     | 25      | 30            | 10         | 50                       | 3000      | 0.1      |
| Bates [6]            | 400     | 10            | 3.5        | 10                       | 2856      | 1        |
| HEPL [41]            | 48      | 0.6           | 50         | 10                       | 11.8      | pulsed   |
| CEBAF-FET [42]       | 45      | 0.3           | 0.2        | 5                        | 1497      | 100      |
| JLab IR Demo [43]    | 20-50   | 5             | 60         | 10                       | 75        | 100      |
| CEBAF-ER [29]        | 1050    | 0.09          | 0.2        | 1                        | 500       | 100      |
| JAERI [44]           | 17      | 40            | 500        | 30                       | 80        | pulsed   |
| BINP [45]            | 22      | 30            | 2000       | 30                       | 22.5      | 100      |
| JLab IR Upgrade [39] | 165     | 9             | 135        | 10                       | 75        | 100      |
| ALICE [46]           | 27.5    | 8.125         | 100        | 1.2                      | 81.25     | 0.1      |
| JLab UV Demo [47]    | 135     | 2.5           | 60         | 5                        | 37.5      | 100      |

Synchrotron light ERLs promise to produce radiation with much enhanced average brightness and peak flux compared to third generation SR sources, very short pulses (1 ps to 100 fs), and high coherence. ERLs also are proposed as part of future electron-ion colliders for nuclear and/or particle physics research [26]. The linac-ring option of Electron-Ion Colliders with target luminosity around  $10^{33} \text{ cm}^{-2}\text{s}^{-1}$ , such as LHeC [36], would solely rely on the ERL concept to limit the total wall-plug power to 100 MW, exploiting the unique features of ERLs. An ERL accelerator combines characteristics of both storage rings and linacs, and is potentially capable of accelerating tens of milliamperes of average current to several tens of GeV. As a result of energy recovery, the rf power required for acceleration becomes nearly independent of the beam current.

#### 4.4.2 System Architecture and Beam Dynamics Issues

Energy recovery is used to mitigate cost in rf- and SRF-driven linacs; economics and performance are dominated by fundamental physical constraints and by management of power flow through the rf drive system (transient management), the beam itself (losses from halo and

coupling of beam to the environment via wakes and impedances), and synchrotron radiation emitted during beam bending. ERLs are transport lines, not closed systems. The beam does not reach equilibrium; beam quality is thus source-limited and dominated by excitation processes (e.g. quantum excitation, wakes, noise etc.). Radiation damping has limited palliative effect; active stabilization (feedback) requires very fast diagnostics and controls. Beam quality and stability are therefore of paramount concern.

ERL dynamics are fully 6-dimensional and benefit from use of emittance exchanges amongst projected subspaces [31]. Though the beam may reside in common regions of configuration space on multiple passes, the energy differs from pass to pass so that (in contrast to storage rings) there is no "closed orbit".

As the system is not cyclic and is of finite length, betatron stability is not inherently required; Twiss parameterizations of the beam (beam betatron functions) cannot, consequently, be uniquely identified with Twiss parameterizations of the lattice transfer map: the beam and the lattice are different. There can in fact be benefits — for the purposes of aberration control — from deliberate mismatch of the beam to the

lattice [32]. The particular scheme used to manage the longitudinal phase space through beam acceleration, delivery, use, and recovery cycle — or “longitudinal match” — is a key feature of any ERL design, inasmuch as it dictates much of the susceptibility of the system to collective effects and defines nearly all features of the rf drive [33] and other aspects of power flow through the system. In this regard, ERLs are essentially equivalent to time-of-flight spectrometers.

The absence of equilibrium, closed orbit, and uniquely defined/equivalent beam/lattice functions elevates halo to the role of a key constraint on ERL performance. High beam power in such systems demands that halo-related losses be limited to a few parts per million of the beam current. However, the nonequilibrium beam behavior renders conventional characterizations of beam and aperture in terms of Gaussian distributions and rms spot sizes (“ $\sigma$ ”) are generally inapplicable [34]. Instead, halo sources must be individually catalogued, their evolution through the system evaluated, and mitigation methods (like collimation, Sec.3.3.11) implemented if needed. Numerous phenomena are of concern in the design and operation of ERLs. These may be categorized as 1) source performance, 2) issues related to beam-quality preservation in high power, high brightness SRF systems, 3) effects in large-acceptance transport systems, and 4) problems related to power-flow management.

**Sources** Performance of ERLs is source limited. An essential element of an ERL is a high performance source capable of production of highest quality beam and then appropriate beam transport to assure emittance preservation in the low energy regime, where the space charge forces have paramount effect on emittance growth. There are presently four major source technologies: SRF, NC rf, dc and thermionic. The use of high accelerating gradients is required to rapidly accelerate the electrons from the cathode to alleviate beam quality degradation. Operating at high repetition rate gives rise to significant thermal effects. Optimization study done at Cornell for a DC gun where space-charge induced emittance growth is most severe has demonstrated emittance compensation for a wide range of bunch charges (0.1 mm-mrad emittance for 0.1 nC charge, 0.7 mm-mrad for 1 nC) [27]. Nuclear and/or particle-physics applications may require a polarized electron source.

### High power, high brightness SRF systems

The use of SRF cavities in high-brightness systems engenders numerous issues associated with beam stability and beam quality preservation. The most evident, historically, is the beam break-up (BBU) instability (Sec. 2.4.2) [25]; recent developments in SRF cavity and accelerator transport system design have rendered this effect largely manageable [35] though means to do so (e.g. H/V phase space exchanges [30]) remain key design features.

Efforts to preserve beam quality during acceleration, delivery, and recovery must be observant of numerous effects driven by interaction of the beam with itself and its environment, some common to storage rings, others more closely associated with linacs and transport systems. Space charge (both transverse (SC) and longitudinal (LSC) (Sec.2.4.4) has significant impact on beam quality early in the acceleration cycle and strongly influences the longitudinal match, limiting the minimum tolerable injection energy and forcing injection of a long bunch with processing to shorter lengths at high energy. Longitudinal matching is also strongly influenced by coherent synchrotron radiation (CSR) (Sec.2.4.15), which degrades beam quality during bunch length compression (Sec.4.5) and increases achievable emittance. As ERLs often compress beams to high peak currents (e.g. for FEL driver operation), wakefield (Sec.2.4.1) and resistive wall effects (Sec.2.4.9) are evident and of concern both for their impact on beam quality and due to localized heating of beam line hardware. Beam and environmental impedances must be carefully characterized and managed.

The combination of LSC, wakes, and CSR can result in microbunching instabilities (e.g. Secs.2.4.15,3.1.7,3.1.8). High energy ERLs with small geometric emittance must (as in high-brightness electron storage rings) contend quantum excitation (incoherent synchrotron radiation) (Secs.3.1,3.1.4.1), which can degrade the beam quality. Preservation of small emittance also forces consideration of intrabeam scattering (Sec.2.4.12), Touschek effect (Sec.2.4.12), and beam-gas scattering (Sec.4.3.1) as sources of emittance degradation and halo formation; the latter can potentially lead to intolerable levels of beam scraping. Given the possible use of high repetition rate with modest or low bunch charges, ion accumulation and trapping (Sec.2.4.13.1) are of concern.

**Accelerator transport** An ERL transport system must meet a number of implicit requirements: it must deliver a properly configured phase space to the user (transverse and longitudinal matching); it must support the decelerating energy-recovery beam; these requirements must be met in a manner assuring beam stability and limiting beam loss to tolerable levels. This generally will require manipulation of phase space to avoid beam loss through the use of appropriate transverse and longitudinal matching.

The use of the beam results in extraction of energy (e.g. via an FEL interaction or emission of synchrotron radiation) with attendant degradation of beam quality. As a result, ERLs often require compression of energy spread during energy recovery to limit losses. Additionally, conservation of energy demands that either the rf drive system compensate for the energy draw by user and collective processes, or that the beam be recovered to a final energy below that at which it was injected, with outcome determined by choice of longitudinal matching scheme. In the event of a severely user-degraded phase space, energy compression requirements may demand that accelerated and recovered beams be less than 180 degrees out of phase with one another, leading to “incomplete energy recovery” [38], with the recovered final energy higher than that at injection. The details of the longitudinal matching process define rf power requirements [33].

Given the potential for large amplitude beam components (halo) and degradation of beam quality during transport and use, ERL transport systems must typically be of large acceptance and incorporate means for the management of aberrations. This is consistent with the view of an ERL as a time-of-flight spectrometer system, and emphasizes that magnetic field quality imposes significant constraints on ERL performance. In particular, field imperfections can lead to energy-spread-inducing phase errors, which will aggravate losses during beam recovery; tolerance to these errors scales inversely with energy, so that the absolute field quality of an ERL must improve as the full energy increases [37].

**Power-flow management** ERLs are intended for use as wall-plug-efficient sources of extremely high power beams (“linac quality beam at storage ring electrical efficiency” [1]); as a consequence the management of power flow via rf drive inputs, beam-environment interactions, beam loss,

radiative processes, and interaction with targets, is a critical aspect of the overall system design and operation so as to avoid beam instability and beam/power-loss induced damage to hardware.

Numerous power-management issues must be addressed. High gradient SRF cavities undergo ponderomotive (or Lorentz force) detuning, which changes under varying beam conditions - such as during beam off/on transients or when a user process (such as an FEL) extracts energy from the beam. The recovery phase - coupled to the energy through the longitudinal match - then shifts, altering the degree of beam loading and shifting cavity resonances. The rf drive must therefore be properly configured to assure stability [33]. The available beam power creates potential for the emission of high power electromagnetic radiation through the coupling of the beam electromagnetic field to the accelerator environment. This of course can be intentional (e.g. ERL-driven FELs), but can lead to unanticipated power loss/deposition through a number of mechanisms. Propagating HOM fields from SRF cavities can convert and transmit substantial power [25], as can the interaction of the beam with any environmental impedance (Secs.3.2.1,2,4.1) – just as in storage rings. Resistive wall effects can, with short bunches and high peak currents, cause significant heating [39].

Coherent synchrotron radiation (CSR) (Secs.2.4.15,3.1.2,3.1.7), can extract extremely large amounts of power from the beam and may result in significant heating of various system components with deleterious effect. This has limited performance in existing oscillator-based ERL-driven FELs [39]. In very high current ERLs operated at high energy, substantial power will be emitted as incoherent synchrotron radiation, at power levels potentially sufficient to warrant engineered controls to avoid heating and hardware damage; the associated degradation of beam quality and loss of energy must be accounted for in the longitudinal matching scheme and design of the recovery process. Use of incomplete energy recovery [38] may be required to adequately manage the longitudinal phase space during deceleration; rf power transients will be at issue as described above.

As beam powers are very high, beam loss must be limited to a small fraction of total beam current (typically a few parts per million) so as to avoid radiation, activation of and/or damage

#### Sec.4.5: BUNCH COMPRESSION

to beamline components. Beam halo is therefore of paramount concern and processes leading to the evolution of halo (field emission in high voltage/rf structures, incomplete extinction ratio in gated photocathode drive lasers, beam dynamical effects during beam formation and handling, intrabeam scattering (Sec.2.4.12), Touschek effect (Sec.2.4.12), ion accumulation (Sec.2.4.13.1)) must be evaluated to analyze beam loss patterns, generate schemes for halo tuning, and/or and provide collimation systems (Sec.3.3.11) for controlled management of beam spill.

#### References

- [1] L. Merminga, D.R. Douglas, G.A. Krafft, Annu. Rev.Nucl.Part.Sci. 53 (2003) 387
- [2] R.E. Rand, Recirculating Electron Accelerators, Accelerators and Storage Rings, Vol. 3, Harwood Academic, New York (1984)
- [3] P. Axel et al, IEEE Tr.NS-24 (1977) 1133
- [4] C.M. Lyneis et al, IEEE Tr.NS-28 (1981) 3445
- [5] H.-D. Graef, A. Richter, LINAC 88, 231
- [6] J. Flanz, C. Sargent, IEEE Trans NS-32 (1985) 3213
- [7] S. Penner et al, IEEE Trans.NS-32 (1985) 2669
- [8] H. Herminghaus et al, IEEE Trans.NS-30 (1983) 3274
- [9] E.A. Antokhin et al, NIM A528 (2004) 15
- [10] C.W. Leemann, D.R. Douglas, G.A. Krafft, Annu. Rev.Nucl.Part.Sci. 51 (2001) 413
- [11] T.I. Smith et al, NIM A259 (1987) 1
- [12] J. Sekutowicz et al, PRST-AB 8, 010701 (2005)
- [13] M. Tigner, Nuovo Cim. 37 (1965) 1228
- [14] G.R. Neil et al, PRL 84 (2000) 662
- [15] R. Hajima et al, NIM A507 (2003) 115
- [16] E.J. Minehara, NIM A557 (2006) 16
- [17] S.M. Gruner et al, RSI 73 (2002) 1402
- [18] Y. Saveliev et al, IPAC 10 (2010) 2350
- [19] K. Umemori et al, SRF-2009 (2009) 896
- [20] I. Sakanaka et al, IPAC 10 (2010) 2338
- [21] V.N. Litvinenko, IPAC 10 (2010) 2364
- [22] J. Flanz, S. Kowalski, C. Sargent, IEEE Trans.NS-28 (1981) 2847
- [23] J. Flanz, PAC 89 (1989) 1349
- [24] V.I. Veksler, USSR Acad. Sci. 43 (1944) 346; J. Phys. USSR 9 (1945) 153
- [25] G.A. Krafft, J.J. Bisognano, PAC 87, 1356
- [26] I. Ben-Zvi et al, ERL 2005 Workshop
- [27] I.V. Bazarov, C.K. Sinclair, PRST-AB 8, 034202 (2005)
- [28] P. Piot, D.R. Douglas, G.A. Krafft, PRST-AB 6, 030702 (2003)
- [29] S.A. Bogacz et al, PAC 03 (2003) 195
- [30] R. Rand, T. Smith, PA 11 (1980) 1
- [31] M. Rihaoui et al, PAC 09 (2009)
- [32] B. Blind, A. Jason, PAC 05 (2005) 225
- [33] T. Powers, C. Tennant, ERL 07, 75
- [34] S. Bernal et al, PRST-AB 14, 104202 (2011)
- [35] C. Tennant et al, PRST-AB 8, 074403 (2005)
- [36] S.A. Bogacz et al, IPAC 11 (2011)
- [37] D. Douglas, BIW 10, 506
- [38] S.V. Benson, D. Douglas, U.S. Patent Number 7,166,973 (2007)
- [39] S.V. Benson et al, PAC 07, 79
- [40] S.O. Schriber et al, PAC 77, 1061
- [41] T.I. Smith, NIM A259 (1987) 1
- [42] N. Sereno et al, PAC 93, 3246
- [43] G. R. Neil et al, PRL 84 (2000) 662
- [44] R. Hajima et al, FEL 06, 312
- [45] O.A. Shevchenko, ERL 11
- [46] Y. Saveliev, ERL 11
- [47] R. Legg et al, PAC 11 (2011)

#### 4.5 BUNCH COMPRESSION

*P. Emma, LBNL*

**Linear compression** Future 4th-generation light sources and linear colliders require very short bunches of high-brightness electron beams. Linac-based free-electron lasers create a strong motivation to compress below the 100-fsec level, which is 2-3 orders of magnitude shorter than in present storage rings. The physics and technical challenges of electron bunch compression are described briefly here.

**RF Acceleration and Energy Chirp** An electron bunch is compressed in length by first accelerating in a linac at an rf phase,  $\phi_{rf}$ , which is not at crest (*i.e.*,  $\phi_{rf} \neq 0$ ), producing a nearly linear energy ‘chirp’ (energy correlation along the bunch),  $h$ , defined as

$$h = -\frac{2\pi eV_0}{\lambda E_f} \sin \phi_{rf} = -\frac{2\pi}{\lambda} \left[ 1 - \frac{E_i}{E_f} \right] \tan \phi_{rf}. \quad (1)$$

Here  $\lambda$  is the rf wavelength and  $V_0$  is the on-crest voltage gain of the linear accelerator, which, at an off-crest phase of  $\phi_{rf}$ , accelerates the electron charge,  $e$ , from an energy of  $E_i$  to  $E_f$ . Signs are chosen such that an rf phase of  $0 < \phi_{rf} < \pi$  sets the head of the bunch ( $z > 0$ ) at a lower energy than the tail ( $h < 0$ ). Longitudinal wakefields of the rf structures typically ‘fight’ this rf-induced chirp, adding an (estimated) opposing chirp of

$$h_w \approx \frac{2NBe^2Z_0cs_0L}{\pi a^2\Delta z^2E_f} [1 - (1 + \sqrt{l_b/s_0})e^{-\sqrt{l_b/s_0}}], \quad (2)$$

where  $l_b$  is the FWHM bunch length (uniform distribution),  $N_B$  is the bunch population,  $Z_0$  is the free-space impedance,  $c$  is the speed of light,  $L$  is the rf structure length,  $a$  is its mean iris radius, and  $s_0$  is the characteristic wakefield length.

Magnetic Compression With an initial rms bunch length of  $\sigma_{z_i}$ , and an initial uncorrelated rms relative energy spread of  $\sigma_{\delta_i}$  at energy  $E_i$ , the final rms bunch length after a magnetic bend system with path length dependence on particle energy, to first order, is

$$\sigma_{z_f} = \sqrt{(1 + hR_{56})^2 \sigma_{z_i}^2 + (E_i/E_f)^2 R_{56}^2 \sigma_{\delta_i}^2}, \quad (3)$$

where  $R_{56}$  is the transfer matrix element which maps relative energy (momentum) deviations,  $\delta \equiv \Delta p/p$ , into longitudinal position deviations within the bunch ( $\Delta z = R_{56}\delta$ ). Full compression is achieved at  $R_{56} = -1/h$ , ‘under-compression’ at  $1 + hR_{56} > 0$ , and ‘over-compression’ at  $1 + hR_{56} < 0$ . A compression factor can be defined as  $C_0 \equiv \sigma_{z_i}/\sigma_{z_f} \approx |1 + hR_{56}|^{-1}$ , with typically  $C_0 > 1$ . The final rms energy spread is also necessarily increased by the linac energy chirp to

$$\sigma_{\delta_f} = \sqrt{h^2 \sigma_{z_i}^2 + (E_i/E_f)^2 \sigma_{\delta_i}^2}. \quad (4)$$

Velocity Compression For a bunch which is not ultra-relativistic, the value of  $R_{56}$  over a drift of length  $L$  at a constant energy  $\gamma = E/mc^2$  is

$$R_{56} = L \frac{1}{\gamma^2}. \quad (5)$$

This can be substituted into  $R_{56}$  of Eq.(3) to calculate the compression over a drift at a constant low energy. It is typically ignorable in high energy linacs, but can be used in the injector [1].

**Types of magnetic compressors** The compression description above assumes a particular  $R_{56}$  value generated by a bending system.

A Magnetic Chicane The simplest, and therefore most common bunch compressor is the chicane, which consists of just 4-dipole magnets and is transversely achromatic to all orders (rectangular magnets), generating an  $R_{56}$  value of

$$R_{56} = 4L_B [\sin(\theta)/\theta - \cos(\theta)] + 2\Delta L \theta \sin(\theta), \quad (6)$$

where  $\pm\theta$  is the bend angle of each dipole magnet,  $\Delta L$  is the drift distance between the 1st and 2nd, as well as the 3rd and 4th magnets, and  $L_B$  is the magnetic length of each dipole, each measured along the beam path with  $E \rightarrow \infty$ . (The distance between the 2nd and 3rd magnets has no

effect on  $R_{56}$ ). The sign convention here has the bunch head at  $z > 0$ , which means a high energy particle passing through a chicane ( $R_{56} > 0$ ) is advanced, arriving early ( $\Delta z > 0$ ).

For  $|\theta| \ll 1$ , the  $R_{56}$  approximates to

$$R_{56} \approx 2\theta^2 \left( \Delta L + \frac{2}{3} L_B \right). \quad (7)$$

This can be substituted into  $R_{56}$  of Eq.(3) to calculate the final bunch length after the chicane.

Bunch compressors also produce non-linear terms, such as:  $\Delta z = R_{56}\delta + T_{566}\delta^2 + U_{5666}\delta^3 + \dots$  In a chicane, their ratios to the linear term are

$$T_{566} \approx -\frac{3}{2}R_{56}, U_{5666} \approx +\frac{4}{2}R_{56}, \text{ etc.} \quad (8)$$

A Simple Dog-Leg Bend A ‘dog-leg’ bend system is composed of two dipole magnets of opposite (or same) sign with quadrupole magnets between and the linear dispersion fully suppressed after the 2nd bend. The total  $R_{56}$  over the system for  $|\theta| \ll 1$  is

$$R_{56} \approx -\theta^2 L_B / 3, \quad (9)$$

where the sign has flipped with respect to a chicane. Unlike the chicane, the higher-order terms are not so simply related, and the  $R_{56}$  is typically quite small compared to a chicane.

A FODO Arc An arc composed of FODO cells is a common compressor with compression matrix element

$$R_{56} \approx -\frac{\theta_T^2 L_T}{4N_c^2 \sin^2(\mu/2)}, \quad (10)$$

where  $\theta_T$  is the net bend angle of the arc,  $L_T$  is the arc length,  $N_c$  is the number of FODO cells, and  $\mu$  is the bend-plane betatron phase advance per cell. The sign is opposite to a chicane and  $T_{566}/R_{56}$  is typically  $> 1$ .

**Nonlinear compression limitations** The effects of an inevitable non-linear energy chirp due to the sinusoidal rf are examined here, along with the second-order compression term,  $T_{566}$ .

Second-Order Limits The second-order chirp,  $h_2$ , due to sinusoidal rf, is defined as

$$h_2 \equiv -\frac{2\pi^2 e V_0}{\lambda^2 E_f} \cos \phi_{rf}. \quad (11)$$

With the further definition  $r \equiv T_{566}/R_{56}$ , the final rms bunch length (squared), to second order, is

$$\sigma_{z_f}^2 = \frac{E_i^2}{E_f^2} R_{56}^2 \sigma_{z_i}^2 + \quad (12)$$

$$(1 + hR_{56})^2 \sigma_{z_i}^2 + R_{56}^2 (h_2 + rh^2)^2 \zeta,$$

where  $\zeta \equiv \langle z_i^4 \rangle - \sigma_{z_i}^4$ , the 4th moment of the distribution. A Gaussian (uniform) distribution produces  $\zeta = 2\sigma_{z_i}^4$  ( $\zeta = 4\sigma_{z_i}^4/5$ ). The term involving  $h_2$  and  $r$  introduces a 2nd-order compression limit, as can wakefields (not included here). If  $h_2 = r = 0$ , then Eq.(12) reduces to Eq.(3).

Second-Order Correction by RF Phasing The second-order aberration will vanish if  $h_2 + rh^2 = 0$ . Equation (11) shows that  $h_2 < 0$  for an accelerating rf phase (*i.e.*,  $-\pi/2 < \phi_{rf} < \pi/2$  and  $eV_0 > 0$ ). Since  $h^2 > 0$ ,  $h_2 + rh^2$  cannot be zero for an accelerating system if  $r < 0$ . Equation (8) shows that a chicane, as well as any other compressor with only dipole magnets, produces  $r < 0$ . Therefore, the second-order compression aberration cannot be canceled with simple rf phasing for an accelerating system ending in a chicane. A decelerating phase can be used, but the energy extracted may make this an unattractive solution. There are, however, other compressors, such as a simple FODO-cell arc which have  $r > 0$ , where it is possible to remove the second-order aberration simply by choosing the right rf accelerating phase, given by

$$\cos \phi_{rf} = \frac{\sqrt{E_i^2 + 8e^2 V_0^2 r(1+2r)} - E_i}{2eV_0(1+2r)} > 0. \quad (13)$$

#### Second-Order Correction by Harmonic RF

Harmonic rf can also be used to compensate 2nd-order limits [2, 3]. The higher frequency section is typically set at the decelerating ‘crest’ phase ( $\phi_n = \pi$ ). The harmonic rf voltage needed for 2nd-order compensation of bunch compression is

$$eV_n \approx \frac{E_f \left[ 1 - \frac{1}{2} r \frac{\lambda^2}{\pi^2 R_{56}^2} (1 - \sigma_{z_f}/\sigma_{z_i})^2 \right] - E_0}{(\lambda/\lambda_n)^2 - 1}, \quad (14)$$

where  $E_0$  is the initial energy before rf acceleration at wavelength  $\lambda$  (*e.g.*,  $E_0 \approx 0$ ), and the harmonic voltage,  $V_n > 0$ , is decelerating for  $r < 0$  and is reduced by the square of the harmonic number ( $n \equiv \lambda/\lambda_n$ ). It is therefore an advantage to use the highest frequency harmonic available (provided that  $2\pi\sigma_{z_i}/\lambda \ll 1$ ). It is also possible to use a harmonic rf phase which is not at the decelerating crest in order to add some linear chirp or even correct third-order correlations. This possibility is not addressed here.

Second-Order Correction by Sextupoles Sextupole magnets can also be used to compensate 2nd-order limits, but this typically requires

several sextupoles and care in balancing the geometric and chromatic aberrations. It is not easily achieved in a chicane.

**Synchrotron radiation effects** An electron beam passing through a bend magnet generates synchrotron radiation, which may degrade the beam brightness.

Incoherent Synchrotron Radiation (ISR) An electron beam of energy  $E$  passing through a bend magnet of length  $L_B$ , bend angle  $\theta$ , and bend radius  $\rho = L_B / \sin \theta$ , generates synchrotron radiation, which adds rms relative energy spread to the beam of [4]

$$\sigma_{\delta_{ISR}} = \sqrt{\frac{55}{24\sqrt{3}} \frac{r_e \hbar c}{(mc^2)^6} \frac{E^5 L_B}{|\rho^3|}} \approx \frac{1}{L_B} \sqrt{(4.13 \times 10^{-11} \text{m}^2 \cdot \text{GeV}^{-5}) E^5 |\rho^3|}. \quad (15)$$

This energy spread is stochastic in character and adds in quadrature to the initial energy spread. A chicane with four dipole magnets, each of length  $L_B$ , will add twice this level,  $2\sigma_{\delta_{ISR}}$ . The energy spread is generated within the chicane and induces an additive bend-plane emittance growth of [5]

$$\Delta\gamma\epsilon \approx (8 \times 10^{-8} \cdot \text{GeV}^{-6}) \cdot E^6 (\Delta L + L_B + \frac{\hat{\beta} + \check{\beta}}{3}), \quad (16)$$

where  $\hat{\beta}$  and  $\check{\beta}$  are the maximum and minimum of the bend-plane beta function within the chicane.

Coherent Synchrotron Radiation (CSR) As the electron bunch is compressed, it may begin to radiate coherently as well, generating further bend-plane emittance growth [6]. A very rough estimate for a Gaussian bunch, including only steady-state longitudinal forces over just the 4th bend of a chicane (where the bunch is shortest in ‘under-compression’), is

$$\left( \frac{\epsilon}{\epsilon_0} \right) \approx \sqrt{1 + \frac{(0.22)^2}{16} \frac{r_e^2 N_B^2}{\gamma \epsilon_N \beta} \psi \left( \frac{|\theta^5| L_B}{\sigma_{z_f}^4} \right)^{2/3}} \quad \psi \equiv L_B^2 (1 + \alpha^2) + 4\beta^2 + 4\alpha\beta L_B, \quad (17)$$

where  $\epsilon_N$  is the initial normalized bend-plane emittance,  $N_B$  is the bunch population, and  $\beta$ ,  $\alpha$  are the nominal bend-plane Twiss parameters at the end of the 4th bend. The effect can be minimized by keeping  $\alpha \approx 0$  and using weak bends, but this formula is too rough to be used to optimize  $\alpha$  and  $\beta$ . Equation (17) is typically an underestimate, since it ignores radiation from the other chicane bends. More information is available in Secs.2.4.15,3.1.7.

**Compressor tolerances** Bunch compression systems may also introduce many tight tolerances, with two such issues highlighted here.

**Compressor Stability** A small rf phase error can produce a large compression error, demanding tight rf phase stability tolerances. The relative compression jitter is related to the rf phase jitter [7]

$$\begin{aligned} \frac{\Delta\sigma_{z_f}}{\sigma_{z_f}} &= -\frac{\Delta C}{C_0} = (C_0 \mp 1)\Delta\phi_{rf} \cot\phi_{rf} \\ &\approx C_0 \frac{\Delta\phi_{rf}}{\phi_{rf}}, \end{aligned} \quad (18)$$

with under-compression expressed by the minus sign and over-compression by the plus sign. The approximation at right is for a high compression factor ( $C_0 \gg 1$ ) and the rf phase not too far off crest ( $|\phi_{rf}| \ll 1$ ) [8].

**Compressor Dipole Field Quality** An energy chirped beam can become very large transversely at the center of a chicane (or similar compressor) where the bend-plane dispersion is largest, requiring demanding dipole field quality tolerances.

The tolerance for quadrupole field component,  $b_1$ , and sextupole field component,  $b_2$ , (both measured at radius  $r_0$ ) in the 2nd and 3rd chicane dipole magnets, each with bend angle  $\theta$ , dispersion  $D$ , bend-plane beta function  $\beta$ , and initial geometric bend-plane emittance  $\epsilon_0$ , expressed as a fraction of the dipole field,  $b_0$ , is

$$\left| \frac{b_1}{b_0} \right| < \frac{1}{|\theta|} \frac{r_0}{\beta} \sqrt{\frac{2\Delta\epsilon/\epsilon_0}{1 + \xi^2}}, \quad (19)$$

$$\left| \frac{b_2}{b_0} \right| < \frac{1}{|\theta|} \frac{r_0^2}{\sqrt{\beta^3\epsilon_0}} \frac{\sqrt{\Delta\epsilon/\epsilon_0}}{1 + \xi^2}, \quad (20)$$

where  $\xi^2 \equiv D^2\sigma_\delta^2/(\beta\epsilon_0)$  and  $\Delta\epsilon/\epsilon_0$  is the tolerable emittance growth (e.g.,  $\lesssim 2\%$ ).

These field errors generate 1st and 2nd order dispersion, which can be corrected with small quadrupole and/or sextupole magnets added to the compressor chicane, if the tight field quality tolerances cannot be met.

## References

- [1] L. Serafini, M. Ferrario, AIP CP 581 (2001) 87
- [2] P. Emma, SLAC-TN-05-004, <http://www.slac.stanford.edu/pubs/slactns/slac-tn-05-004.html>
- [3] F. Zimmermann, T.O. Raubenheimer, AIP Conf.Proc.367 (1996) 84
- [4] M. Sands, SLAC-121 (1970)
- [5] T.O. Raubenheimer, P. Emma, S. Kheifets, PAC 93, 635
- [6] E.L. Saldin, E. A. Schneidmiller, M.V. Yurkov, NIM A483 (2002) 516
- [7] K.L. Bane et al, PAC 93, 596
- [8] M. Dohlus, T. Limberg, FEL 05, Stanford

## 4.6 OPERATION OF LINEAR-COLLIDER FINAL FOCUS SYSTEMS

A.A. Seryi, JAI, G.R. White, SLAC

The Final Focus System (FFS) is part of the Beam Delivery System (BDS) which is responsible for transporting the  $e^+/e^-$  beams from the exit of the high energy linacs, focusing them to the sizes required to meet the luminosity goals, bringing them into collision, and transporting the spent beams to the main beam dumps. In addition, the BDS must measure the linac beam and match it into the FFS; protect the beamline and detector against mis-steered beams from the main linacs; remove any large amplitude particles (beam-halo) from the linac to minimize background in the detectors; measure and monitor the key physics parameters such as energy and polarization before and after the collisions. The BDS must provide sufficient instrumentation, diagnostics and feedback systems to achieve these goals.

The main subsystems of the BDS needed to fulfill the listed goals are the diagnostics region, fast extraction and tuneup beamline, betatron and energy collimation, FFS proper, interaction region (IR) and extraction line. The main task of a FF is to focus the beams to the small sizes required at the interaction point (IP). To achieve this, the FFS forms a large and almost parallel beam at the entrance to the Final Doublet (FD), which contains two or more strong quadrupole lenses. The high energy linear colliders require  $\sigma_x \gg \sigma_y$  at the IP in order to maximize the luminosity while handling the beam-beam interaction. Typical designs have a demagnification of a few hundred in  $y$  with an order-of-magnitude less in  $x$ . The  $y$ -focusing is usually set close to the limit from the depth of focus, i.e.  $\beta_y^* \sim \sigma_z$ , which can be extended by a factor of several by travelling focusing, but other constraints, e.g. chromaticity or tolerances may limit the achievable depth of focus.

Most next generation linear colliders have beam collisions with a crossing angle  $\Theta_c$  to avoid

## Sec.4.6: OPERATION OF LINEAR-COLLIDER FINAL FOCUS SYSTEMS

parasitic bunch collisions and to provide a method of directing the spent beams to dumps. The crossing-angle will reduce the luminosity unless the beams are “crabbed”, i.e. tilted in  $x - z$  plane so that they pass through each other head-on.

Significant advances in the optical design and chromatic correction of the linear-collider FFS have been made since described in earlier editions of this handbook [1].

### 4.6.1 Chromaticity Compensation

Even for a beam with a minor energy spread of a fraction of a percent, the focused beam size will be diluted by the chromaticity of these strong lenses. The design of a FF is therefore driven primarily by the necessity of compensating the chromaticity of the FD.

There are two primary approaches for chromaticity compensation — the non-local scheme, implemented particularly at FFTB [2] and B factories [3, 4] and the local compensation scheme [5] presently being implemented at ATF2 [6], which is used as a basis for the FFS design of proposed linear colliders.

In the non-local FFS the chromaticity is compensated in dedicated chromatic correction sections by sextupole magnets placed at maxima of dispersion and beta-functions. The geometric aberrations generated by the sextupoles are canceled by using them in pairs with a minus identity transformation between them. The non-local FFS is built from separated optics blocks with strictly defined functions, and is relatively simple for design and analysis. The major challenge of the non-local FFS is in its applicability to high energy colliders — the FFS bend magnets have to be sufficiently long and weak to minimize the additional energy spread generated, lengthening the system to the range of several kilometers which makes it impractical to use it for TeV scale  $e^+e^-$  colliders.

Fig.1 shows conceptual optics layout of FFS with local chromaticity compensation. Local

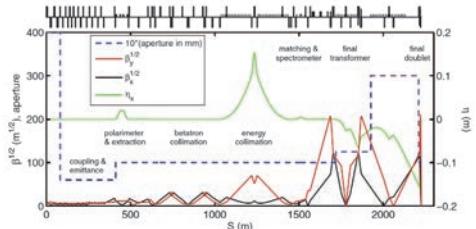


Figure 2: Optics of ILC beam delivery system [7].

compensation of chromaticity is achieved by interleaving a pair of sextupole magnets with the quadrupoles of the final doublet. The dispersion throughout the FD is created by upstream bends, and is designed to be zero at the IP. Geometric aberrations, generated by FD sextupoles are cancelled by two or more sextupoles located upstream. Sextupoles placed in FD generate second order dispersion, which, however, can be compensated simultaneously with  $x$  and  $y$  chromaticity provided that half of the total horizontal chromaticity of the whole final focus is generated upstream. The optics of the ILC Beam Delivery System [7] was designed according to this described approach aiming for up to 1 TeV CM. This is shown in Fig.2.

**Luminosity** The luminosity is

$$\mathcal{L} = \mathcal{L}_0 H_D \quad (1)$$

where  $H_D$  is the luminosity enhancement (Sec.2.5.3),  $\mathcal{L}_0$  is the geometric luminosity which depends on the beam properties as well as the FFS properties such as the depth of focus and the *bunch* crossing angle  $\theta_c$  ( $\theta_c = 0$  with crab crossing;  $\theta_c = \Theta_c$  without crab crossing),

$$\mathcal{L}_0 = \frac{f_{\text{rep}} n_b N_b^2}{4\pi \sigma_{x,y}^*} \eta(C_\theta, A_y) \quad (2)$$

where  $f_{\text{rep}}$ ,  $n_b$ , and  $N_b$  are the repetition rate, the number of bunches per pulse, and the number of particles per bunch,  $\sigma_{x,y}^*$  are the rms beam sizes at the IP and function  $\eta$  describes the effects due to depth of focus and the crossing angle. Assuming flat Gaussian beams ( $\sigma_x \gg \sigma_y$ ),  $\eta(C_\theta, A_y)$  is given by the quantity  $R_L$  of Eq.(6), Sec.4.1. The geometric luminosity loss is 14% when  $A_y = \sigma_z/\beta_y^* = 1$  and another 10% when  $C_\theta = \theta_c \sigma_z / \sigma_x = 1$ . Typically, FFSs are designed with  $A_y \lesssim 1$  and  $C_\theta \lesssim 1$ . Crab crossing and “travelling focus” [8] have been proposed to reduce the sensitivity to the crossing-angle and to

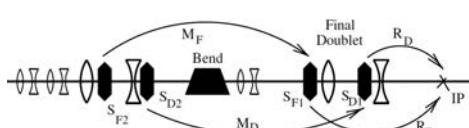


Figure 1: Schematic FFS with local chromatic correction.

the depth of focus, respectively, and are being implemented in the FFS design.

**Crossing-angle and crab-crossing** We address four issues with respect to a crossing angle:

(i) **Geometry** If  $\Theta_c$  is large, it may be possible to have separate quads for the incoming and outgoing beams while for small  $\Theta_c$  an exit port is usually needed in the FD quads. In all cases,  $\Theta_c$  needs to be sufficiently large to separate the beams and to allow them to escape the IR.

(ii) **Multibunch crossing instability** [9] An instability can arise from the parasitic bunch interactions. Assuming flat beams, the criterion to avoid significant vertical displacement of the bunches due to this instability is

$$(m_B - 1) \ll \frac{C_\Theta^2}{\mathcal{D}_x \mathcal{D}_y} \sqrt{\frac{1}{2} + \frac{\mathcal{D}_y}{3}} \quad (3)$$

where  $\mathcal{D}_{x,y}$  are the beam-beam disruption parameters (Sec.2.5.3),  $C_\Theta = \Theta_c \sigma_z / \sigma_x$ ,  $m_B$  is the number of bunches that interact parasitically. Assuming no additional masking,  $m_B = 1 + 2L^*/S_B$  with  $S_B$  the bunch spacing and  $L^*$  the free distance to the IP. This instability decreases rapidly with larger  $\Theta_c$ .

(iii) **Crab crossing** When  $C_\Theta \gtrsim 1$ , the beam can be crabbed [10]. Here, the bunches are given a  $z$ -dependent offset so that they collide head-on:  $dx^*/dz = \Theta_c/2$ . There are two approaches: either use an rf deflecting cavity or use the correlated energy spread along the bunch and a small amount of residual dispersion  $D_x^*$  at the IP [11]. The required  $D_x^*$  is

$$D_x^* \frac{\sigma_{\delta_c}}{\sigma_z} = \Theta_c/2 \quad (4)$$

where  $\sigma_{\delta_c}$  is the correlated component of the rms relative energy spread, while the required rf voltage is

$$eV_{rf} = \frac{\Theta_c E \lambda_{rf}}{4\pi R_{12}} \quad (5)$$

where  $E$  is the beam energy and  $R_{12}$  is the transport matrix element (Sec.2.2.1) from the crab cavity to the IP.

Tolerances on the alignment and stability of the crab cavity tend to be loose except for the relative phase tolerance between the crab cavities on either side of the IP. This phase difference  $\Delta\phi_{rf}$  can cause an  $x$  offset between the beams. To limit the luminosity dilution to 2%, the bunch separation must be  $< 0.3\sigma_x$  rms, which in turn requires

$$\Delta\phi_{rf} < 2\pi \frac{0.6\sigma_x}{\Theta_c \lambda_{rf}} \text{ [radians]} \quad (6)$$

The tolerance on the bunch timing is much looser and just shifts the longitudinal IP position with a luminosity loss that depends on the depth of focus.

(iv) **Solenoidal field effects** With a crossing angle, the beams do not travel parallel to the solenoidal field. This results in  $y$  deflections of the beams and vertical dispersion at the IP. Assuming the FD quads are outside of the solenoidal field, the offset is

$$\Delta y^* = \frac{B_s L^{*2}}{B\rho} \frac{\Theta_c}{4} \quad (7)$$

where  $B_s$  is the solenoidal field. This deflection can be corrected by steering but it is more difficult to correct the associated dispersion. There are additional effects arising in the practical case when solenoidal field is partly overlapping with the FD, which is discussed below.

**Aberrations** In the FFS with local compensation of chromaticity, the second order aberrations are canceled when the  $x$  and  $y$  pairs of sextupoles are separated by transfer matrices  $M_{F,D}$  which have block-diagonal structure

$$M = \begin{pmatrix} F & 0 & 0 \\ F_{12} & 1/F & 0 & 0 \\ 0 & 0 & F & 0 \\ 0 & 0 & F_{34} & 1/F \end{pmatrix}, \quad (8)$$

where all non-zero elements are arbitrary. In order to cancel the second order chromatic aberrations the sextupole integrated strength has to satisfy [5]:

$$K_{SF2} = -F^3 K_{SF1}; K_{SD2} = -D^3 K_{SD1}$$

$$K_{SF1} = \frac{\xi_{x1} + \xi_{x2}}{R_{F12}^3 D'}; K_{SD1} = \frac{\xi_y}{R_{D34}^3 D'_{IP}},$$

where  $\xi_{x1}$  is the horizontal chromaticity of the system upstream of the bend,  $\xi_{x2}$  is the chromaticity downstream,  $\xi_y$  is the vertical chromaticity,  $D'_{IP}$  is the angular dispersion at the IP and the transfer matrices are as shown in the figure above. The final-focus chromaticity  $\xi$  is defined as integral over  $s$  of the product of quadrupole strength  $K_1$  and beta function, without any factor of  $4\pi$ . The condition  $\xi_{x1} = \xi_{x2}$  means that half of the horizontal chromaticity is generated upstream, allowing simultaneous compensation of the  $x$  and  $y$  chromaticity together with the second order dispersion. The flexibility to modify the coefficients in Eq.(9) via a proper choice of optics allows compensation of third and fourth order aberrations. Further improvements can be achieved with an additional bend upstream of the second sextupole pair, a further sextupole and an additional decapole.

**Synchrotron radiation** The FFS with local compensation of chromaticity is more amenable for its scaling into the multi TeV region of energies. For a wide range of parameters, the IP beam size dilution for this FFS is dominated by the energy loss due to synchrotron radiation in the bends (this additional rms energy spread can be obtained using Eq.(17), Sec.3.1.4), which scales as [5]:

$$\frac{\Delta\sigma_y^2}{\sigma_y} \propto \frac{\gamma^5}{L^2} D'_B^{3/2} \propto (\gamma\epsilon_y)^{3/2} L^{*3} \left( \frac{D'_{IP}}{\epsilon_x} \right)^{3/2} \left( \frac{\epsilon_x}{\epsilon_y} \right)^{3/2} \frac{\gamma^{7/2}}{L^5}. \quad (9)$$

where  $D'_B$  is the angular dispersion produced by the bend magnets which scales as  $D_B \sim D'_{IP} L^*/L$  and the bend length is assumed to be proportional to the total length of the system  $L$ . The last two terms in parentheses are constant if  $D'_{IP}$  is proportional to the horizontal beam divergence and if the ratio between horizontal and vertical emittance is constant with energy. We also assume constant IP  $\beta$ -functions and  $x$  and  $y$  chromaticity, which is equivalent to holding  $L^*$  and the length of the final quadrupoles fixed. This implies that the scaling is valid only in the range where the gradients required can be provided by increased pole-tip field or decreased aperture. If the normalized emittance stays constant, which is a worst case assumption, then the length of the FFS system scales with energy as  $L \sim \gamma^{7/10}$ . If however the achievable normalized emittance decreases inversely with energy as is assumed in [12], then the length of the FFS scales as  $L \gamma^{2/5}$ . In this case, it is possible to design a FFS for a 3 TeV center of mass energy collider in a length of about 600 m.

Another effect due to synchrotron radiation (Oide effect) is discussed in Sec.3.3.5.

**Effects of the detector solenoid** The field of the detector solenoid produces beam coupling and, when including a crossing angle, it also generates additional dispersion and vertical trajectory through the IP region. The latter will also result in additional IP beam size growth from synchrotron radiation.

In the ideal case when the solenoid field does not overlap with the FD quadrupoles, the symmetry is preserved, and correspondingly various coupling terms are exactly zeroed. The vertical orbit deviation reaches  $\delta y / B_0 L^{*2} \theta_c / (B\rho) / (32)$  at the maximum deviation, however will be exactly

zero at the IP. The beam will reach the IP with the vertical angle  $\delta y' \sim B_0 L^* \theta_c / (B\rho) / 8$ , antisymmetrically with the opposite beam (for e+/e- case) resulting in zero overall crossing angle in vertical plane.

The more realistic case when the detector solenoid field overlaps with the FD means that the symmetry is broken, resulting in anomalous coupling and dispersion created at the IP, as well as non-zero vertical offset of the beam at the IP. The effects can be calculated exactly [13], and can be estimated as

$$y_{IP} \approx \frac{Bl}{B\rho} \frac{\theta_c}{2} L^*; \quad <yx'> \approx \frac{\sigma_{xpo}}{\sigma_{yo}} \frac{BL}{B\rho} L^*; \\ <yE> \approx \frac{\sigma_E}{\sigma_{yo}} \frac{Bl}{B\rho} D'_{x0} L^*. \quad (10)$$

These effects need to be compensated and this can be done most efficiently with use of weak antisoleneoids. Such compensation removes the anomalous effects in the entire energy range of the collider.

The detector solenoid and resulting vertical curved orbit produce synchrotron radiation generated energy spread and corresponding growth of the beam size at IP

$$(\delta\sigma_y^{SR})^2 = C_E \gamma^5 \int R_{36}^2(z) \frac{dz}{|\rho(s)|^3}, \quad (11)$$

where  $C_E = 55r_e \lambda_e / (3^{1/2} 24)$ . The beam size growth is proportional to  $(B_0 L \theta_c)^{5/2}$  and does not depend on the beam energy. However, the nominal beam size typically decreases with energy, and therefore this SR beam size growth must be kept small enough not to limit performance at higher energies.

The Detector Integrated Dipole (DID) is a coil integrated into the detector solenoid, producing sine-like transverse field through the detector, allowing compensation of the angle at the IP and minimization of the vertical orbit variation (thus also the reduction of the synchrotron radiation effects) [14]. A negative polarity of DID is also considered in order to reduce the effect of background due to beam-beam pairs — in this case the anti-DID field is optimized in such a way that most of low energy pairs are guided into the exit aperture of the extraction line [15].

**Wakefields** Wakefields are dangerous in the FFS because the  $\beta$ -functions are large in places and due to the necessity to install collimators and protection masks close to the beam.

Wakefields from the transverse geometric and resistive wall can be important through two effects: (a) increase of the beam emittance, and (b) amplification of the centroid jitter. To fully investigate the impact of component and geometry choices in the FFS, calculated wakes are typically inserted into beam tracking codes. [16], [17] and [18] describe analyses investigating: geometric and resistive wall wakefields of the tapered and flat parts of collimators; resistive wall wakes of the beam pipe, which are especially important in the regions of the final quadrupoles (where the  $\beta$  functions are very large); trapped modes in crab cavities (which can be used to tilt the bunches at the IP to compensate for a crossing angle).

Effect (b) often imposes the more restrictive constraint. The centroid jitter amplification due to the resistive wall wakefield is

$$\frac{\Delta y^*}{\sigma_y^*} = 1.6 \frac{N_b r_e}{\gamma \sigma_z} \frac{L \beta_{FD} \delta_s}{b^3} \left( \frac{\Delta y}{\sigma_y} \right)_{FD} \quad (12)$$

with  $L$  the length of the aperture,  $b$  the beam-pipe radius,  $\beta_{FD}$  the  $\beta$ -function in the FD,  $\delta_s = \sqrt{\sigma_z/Z_0 \sigma_c}$  skin depth.

The effect due to the discontinuity of a gentle taper is [19]

$$\frac{\Delta y^*}{\sigma_y^*} = 0.56 \frac{N_b r_e}{\gamma \sigma_z} \frac{L \Theta^2 \beta_{FD}}{b_1 b_2} \left( \frac{\Delta y}{\sigma_y} \right)_{FD} \quad (13)$$

where  $L$  and  $\Theta$  are the length and angle of the taper,  $b_1$  and  $b_2$  are the initial and final beam-pipe radii. Both these tolerances can be minimized by enlarging  $b$ .

## Scattering

Bremsstrahlung The cross-section for bremsstrahlung at high energies (assuming complete screening) is [20]

$$\begin{aligned} \sigma_{Br} \approx & \frac{16}{3} \alpha Z (Z + 1.35) r_e^2 \ln \left( \frac{183}{Z^{1/3}} \right) \\ & \times \left[ \ln \left( \frac{\delta_{max}}{\delta_{min}} \right) + \delta_{min} - \delta_{max} \right] \end{aligned} \quad (14)$$

where  $\delta_{min}$  and  $\delta_{max}$  are the minimum and maximum photon energy in units of the beam energy. The factor  $Z(Z + 1.35)$  accounts for the nuclear charge and approximates the atomic electrons. Typical cross-section for CO with large energy losses  $\sim$  a few barns. The number of scattered beam particles is  $\Delta N/N = n_{gas} L \sigma_{Br}$ , with  $L$  the distance the beam travels,  $n_{gas} [\text{m}^{-3}] = 3.2 \times 10^{22} P [\text{Torr}] N_{atom}$  at 300° K,  $N_{atom}$  = the number of atoms per gas molecule.

Thermal photons The inverse-Compton scattering with thermal photons is discussed in Sec.3.3.4. To calculate the number of large amplitude scatterings, the differential cross-section needs to be integrated numerically. However, an estimate can be obtained over a distance  $L$ :

$$\frac{\Delta N}{N} = 0.5 n L \sigma_c \quad (15)$$

where  $\Delta E/E = x_{ave}/(1 + x_{ave})$  is the energy deviation of the particle due to scattering, and  $\sigma_c$  is given by substituting  $x$  by  $x_{ave} = 10.8 E k_B T / (m_e^2 c^4)$  in Eqs.(5-6), Sec.3.3.4. For typical parameters, Eq.(15) tends to overestimate the number of large amplitude scatterings by a factor of 2-5.

Elastic Coulomb scattering off nucleus The energy change of the incident particle is relatively small and the primary effect is an angular deflection that may cause the particle to exceed the beam-pipe aperture. The differential cross-section for Coulomb scattering on atomic nuclei is found in Eq.(2), Sec.3.3.1. The neglect of the nuclear form factor and the recoil of nucleus causes a slight overestimate of the particle loss.

Assuming the aperture is limited at a single location by  $b_{ap}$ , such as the final doublet, the number of particles scattered to an amplitude greater than the aperture over a distance  $L$  is

$$\frac{\Delta N}{N} = n_{gas} L \frac{2\pi Z^2 r_e^2}{\gamma^2 b_{ap}^2} \int ds (R_{12}^2 + R_{34}^2) \quad (16)$$

where  $R_{12}, R_{34}$  are from the scattering position to the aperture, and we assume  $b_{ap}^2 \gg (R_{12}^2 + R_{34}^2) \theta_{min}^2$ .

Elastic Coulomb scattering off atomic e<sup>-</sup> [21] The angular deflection can be accounted for by replacing  $Z^2$  with  $Z(Z + 1)$  in Eq.(16). However, the recoil of the electron results in a significant energy change to the incident particle. The differential cross-section for a relative energy change of  $\delta$  is

$$\frac{d\sigma_{ee}}{d\delta} = \frac{2\pi Z r_e^2}{\gamma} \frac{1}{\delta^2} \quad (17)$$

and the number of particles scattered beyond a limiting energy aperture  $\delta_{min}$  is

$$\frac{\Delta N}{N} = n_{gas} L \frac{2\pi Z r_e^2}{\gamma} \frac{1}{\delta_{min}} \quad (18)$$

**Tolerances** The “bare” tolerances one can define for elements of the FFS are extremely restrictive, there is no expectation that real-world alignment and mechanical fabrication processes can be

controlled well enough to suffice. One designs tuning knobs to specifically remove the expected aberrations produced. For all classes of elements in the FFS, we can assign two different types of tolerance.

*Capture tolerances* define the tolerances to which the elements must be tuned before global tuning knobs are applied, in order to ensure that the knobs will reduce the IP spot size / obtain required luminosity within a finite number of iterations. These tolerances strongly depend upon the specifics of the hardware and algorithmic implementation of the global tuning knobs, and are thus hard to specify in a general sense. For example, tuning knobs for ATF2 require coordinated movements of the FFS sextupole magnets to introduce known aberrations at the IP. The range of motion of the magnets is limited by the physical range of motion of the cam-based magnet mover system. This provides a capture range for specific IP aberrations that the mover-based tuning system can deal with, which in turn provides capture tolerances for FFS elements which produce these aberrations. Also, imperfections in the motions of the magnets, field errors in the magnets and non-linearities of the knobs themselves mean that the knobs are not perfectly orthogonal over their entire range. This provides for a soft capture tolerance that is hard to uniquely define.

*Stability tolerances* specify how far a given element may drift from its post-tuneup state before enlarging the IP beam size by a given amount. These are easier to define in generic terms.

We consider here effects that contribute to the vertical IP spot size.

Incoming beam If the geometric aberrations in the FFS are well corrected, the nonlinear aberrations will not be very sensitive to incoming betatron oscillations or changes in the incoming beam emittance. Provided the design optical transformations between the FFS sextupoles exist, these will cause most aberrations to cancel. However, the Oide effect can impose constraints on the incoming trajectory jitter.

Trajectory errors The dominant source of  $y$  trajectory errors are movements of the quads  $y_q$  and rolls of the bends  $\theta_b$ . The position at the IP is

$$\frac{\Delta y^*}{\sigma_y^*} = \int ds R_{34} \left\{ \begin{pmatrix} y_q k_1 \\ \theta_b/\rho \end{pmatrix} \right\} \quad (19)$$

$$= \left\{ \begin{pmatrix} y_q k_1 \\ \theta_b \Theta_b \end{pmatrix} \right\} \sqrt{\frac{\beta_y}{\epsilon_y}} |\sin(\psi_y^* - \psi_y)| \quad x$$

where  $\Theta_b$  is the bending angle of the bend,  $\beta_y$  is evaluated at the magnet, and  $k_1$  is the integrated quadrupole strength  $K_1$ . For most magnets in the FFS,  $|\sin(\psi_y^* - \psi_y)| \approx 1$ .

The tightest sensitivity is usually that on the final doublet. When it moves as a unit, the trajectory motion at the IP is roughly equal to the motion of the magnets. The sensitivity is even tighter when the F and D magnets move in opposite directions. Fortunately, ground motion tends to be correlated over long distances, and the sensitivities are greatly reduced [22]. Also, for a linear collider operating in multi-bunch mode, there will always be the requirement to have a fast feedback system operating at the IP which also greatly relaxes this sensitivity [23].

Dispersion Dispersion generated by deflections (in a FFS with locally corrected chromaticity) is mostly due to the chromatic dependence of the deflection rather than any chromatic dependence of the displaced downstream trajectory,

$$\Delta\sigma_y^* = \left\{ \begin{pmatrix} y_q k_1 \\ \theta_b \Theta_b \end{pmatrix} \right\} \sigma_\delta |R_{34}| \quad (20)$$

Vertical dispersion can also be generated by rolls of the quads  $\theta_q$  or displacements of the sextupoles  $y_s$ ,

$$\Delta\sigma_y^* = \left\{ \begin{pmatrix} 2\theta_q D_x k_1 \\ y_s D_x k_2 \end{pmatrix} \right\} \sigma_\delta |R_{34}| \quad (21)$$

where  $D_x$  is the horizontal dispersion evaluated at the magnet.

Skew coupling Skew coupling is generated by direct skew fields or by trajectory offsets in sextupoles. The direct skew fields can arise from rolls of the quads or displacements of the sextupoles, and from any detector solenoid.

$$\Delta\sigma_y^* = \left\{ \begin{pmatrix} 2\theta_q k_1 \\ y_s k_2 \end{pmatrix} \right\} \sigma_x |R_{34}| \quad (22)$$

where  $\sigma_x$  is the  $x$  beam size at the magnet. The tightest sensitivities are usually on the FD magnets. The FFS must be designed with tuning elements to correct for the skew coupling as part of the design optics including the detector solenoid. With flat beams, the skew correction only needs to be applied to the two phases that enlarge the vertical beam size at the IP, i.e. the  $x^*-y^*$  and  $x'^*-y^*$  terms. As most of the FFS elements are at a phase  $90^\circ$  from the IP, the dominant correlation term to be corrected is usually the  $\langle x'y \rangle$  term.

**Magnet field quality** Unwanted higher-order multipoles in the magnets of the FFS can be a problem due to the large beta functions in parts of the system. This was found to be especially true at ATF2 for example, where the relatively low beam energy (1.3 GeV) means a much larger beam size to magnet aperture diameter ratio compared to the high-energy LC designs that it is testing. Normal and skew multipole tolerances were set based on IP beam size growth estimates from particle tracking codes such as described in the section on optimization and tuning [24]. Skew multipole sensitivities are the most restrictive for vertical beam size at the IP, especially skew sextupole. It is possible to ameliorate the tightness of this sensitivity given precise measurements of the multipole strengths and angles. This is done through non-linear retuning techniques using the FFS sextupole strengths and matching quads and/or increasing  $\beta_x^*$ . The latter reduces the horizontal beam size in the magnet apertures, which causes less coupling into the vertical plane at the IP.

**Magnet field strength** Initial errors in the field strength of the FFS magnets, drifts of field strength over time due to temperature variation or power supply setting errors. These cause direct growth of the vertical beam size at the IP due to introduced mismatches from the design optics, especially in the sextupoles which perform fine cancellations of chromatic, geometric and chromo-geometric terms. They also cause errors in the tuning knobs that rely on the manipulation of those magnets to generate orthogonal sets of aberrations at the IP. Typically, the magnet fields need to be accurate at the 1e-4 to 1e-3 level (dB/B). Some magnets are more sensitive than others; the strong FD magnets, magnets at locations with high beta functions and the sextupoles are especially sensitive.

**Optimization and tuning** The design of the compact FFS optics, and any optimization of that design, for example to change the focal length, or demagnification of the system, or energy bandwidth etc, requires many free parameters to be tweaked in order to control the cancellation of various important aberrations. Specific automatic design and analysis tools have been available in the past to handle generic FFS's, such as FFADA [25]. The optimization of the compact FFS however is much more tricky and only semi-automated tools exist for this task. Published “recipes” exist which describe the design

procedures, but these should be considered guides more than a set of detailed steps, and require a certain amount of intuition in their application. For example, the recipe for optimizing the ILC FFS design [26] uses a semi-automatic design tool based on Matlab together with MAD [27] for basic matching and a combination of TRANSPORT and tracking codes such as TURTLE [28] to determine third and fourth order matrices and calculate beam spot sizes at the IP. As one tries to push the demagnification of the system to its ultimate, and aim for very small spot sizes  $\leq 1$  nm, for example as required by the CLIC [29] design, then one needs to take into account even higher order aberrations. The recipe in [30] deals with aberrations up to ninth-order. Here a code based on PYTHON, MAPCLASS [31] uses MADX [32] together with the polymorphic tracking code PTC [33] to compute transfer maps up to arbitrary orders across the FFS beamline. Tracking codes such as TURTLE, PLACET [34], DIMAD [35], Lucretia [36] etc are required to determine the beamsize, GUINEA-PIG [37] is used to determine the luminosity for the non-Gaussian beam profiles at the IP. The tracking code is also essential for the higher demagnification systems to include the effects of synchrotron radiation. These recipes require much back-and-forth between the various codes to determine the aberrations, beam sizes and to create and use complex knobs to perform the optimizations. This is why to help automate the process an intermediary scripting language such as PYTHON or Matlab are used, both to handle the optimization process and to transfer the lattice and changes thereof between the various codes.

**Online tuning of the FFS** [38] As discussed in the tolerance section, the tolerances of the FFS are usually sufficiently tight that there is no hope that magnet manufacturing and alignment procedures can be good enough by themselves. It is essential to design online tuning procedures that use live, beam-based information to perform the final tuning of the system to achieve the design beam parameters at the IP.

We assume here an incoming matched, dispersion and coupling corrected beam using the upstream BDS optics, such that the remaining aberrations are within the capture tolerances of the FFS tuning knobs. The tuning of the FFS starts with the traditional tasks of beam-based alignment (BBA), aligning the electrical readouts

of the BPMs with the magnet field centers of the magnets they are attached to [39], and using global steering algorithms [40] to put the beam in a straight trajectory through the system. Using beam tracking code, we model the FFS optics, including expected error sources and magnitudes and the expected performance of the BBA process. This modeling is done in a Monte Carlo fashion, each seed of the Monte Carlo simulation contains a different spread of possible errors using a distribution based upon expectations from the various hardware systems. Where possible, measurements and direct experience from beam operation are included. The outcome of this modeling is to derive a list of the most likely aberrations at the IP. We then design a set of tuning knobs to remove these. The knobs are designed to each orthogonally target one of the expected aberrations. To test how robust these knobs are likely to be in the face of a beamline with realistic errors, the tuning procedure is performed in simulation using the Monte Carlo model developed for this. The basis for the tuning knobs is the deliberate offset of the chromaticity correction sextupole magnets. Horizontal and vertical translations, together with roll about the beam axis DOF of these sextupoles generates different sets of first and second-order aberrations at the IP which are calculated to form a response matrix. This is then inverted to provide a set of approximately orthonormal knobs to correct the desired list of aberrations. The simulation is then used to estimate a sequence of knobs to use to remove all important IP aberrations. The most dominating terms are usually the vertical dispersion, waist shift and  $\langle x'y \rangle$  coupling terms. Horizontal and vertical motion of the sextupoles generates these terms as follows (considering the more sensitive vertical dimension here).

Horizontal offset of a sextupole produces a feed-down normal quad field, which can be used to generate a waist shift at the IP by,

$$\Delta\alpha_y^* \approx -K_2 L \Delta x \beta_x^* \beta_y^* \cos 2\Delta\mu_y^s \quad (23)$$

where  $K_2$  is the sextupole strength,  $s$  denotes the location of the sextupole and  $*$  the IP.

Vertical motion of the sextupoles generates a feed-down skew-quad field component that is used to couple the x-y motion of the beam particles, and in areas of design horizontal dispersion additionally produce vertical dispersion.

$$\Delta D_y^* = -K_2 L \Delta y D_x^2 \sqrt{\beta_x^* \beta_y^*} \sin \Delta\mu_y^s \quad (24)$$

It is noted from simulation studies that the tuning of the FFS beamline in the presence of

realistic errors becomes more difficult the more aggressively one tries to focus the beam at the IP; ie the difficulty scales with the chromaticity of the system. For systems with very small IP spot sizes such as required by CLIC, one needs a more non-linear approach to tuning making additional use of octupole and decapole magnets to correct for the more dominant higher-order aberrations during the latter stages of tuning.

**Operational aspects** In order to operationally test the design theories behind the compact FFS optics, the test accelerator ATF2 [41] was constructed. ATF2 has a scaled version of the ILC FFS optics designed to accept an extracted 1.3 GeV electron beam from the ATF damping ring with 12 pm vertical emittance and focus it down to 35 nm vertically at the IP. Configured in this way, the ATF2 FFS optics has a level of chromaticity equal to that of the ILC design and provides a proof of concept test of its operation. There is also the possibility to further squeeze the beta functions at the IP to produce a vertical beam size down to about 22 nm that would test our ability to tune a more highly chromatic system more relevant to a machine such as CLIC. At the time of writing, the accelerator is still being commissioned; a summary of progress so far is available [42].

In addition to testing the optics itself, ATF2 provides a test bed for the various high-precision diagnostics that are required for a future Linear Collider to operate. These are such devices as high resolution cavity BPMs (<100 nm resolution), laserwire systems for non-destructive beam size measurements, fast feedback systems for multi-bunch orbit control. It is not enough just to deliver a small beam to the IP to achieve high luminosity, the beam at the IP must remain in collision with its colliding partner beam at a level small compared to the beam size itself. This typically means a tolerance of the IP beam position at the sub-nm level. Such stability is only possible through a set of low and high bandwidth orbit feedback systems. A secondary aim of the ATF2 project is to demonstrate the required stability required for the maintainable, high luminosity operation of the FFS by demonstrating both a continuous small beam spot size at the IP and holding it stable at the few-nm level.

Simulations of the ATF2 system with expected levels of mechanical vibration and ground motion show an expected growth rate of the beam

size at the IP of about 0.5nm per hour. In addition to holding the orbit steady with beam feedbacks, this means we also have the requirement to periodically re-tune the beam using our developed tuning knobs. It is envisioned that we will apply these using a knob dither-feedback system.

**Machine Detector Interface** The MDI design aimed to provide the necessary luminosity while also providing satisfactory background environment, and beam instrumentation at the IP including the fast luminosity monitor, as well as beam position monitors required for the intra-train feedback. The ILC luminosity can be measured with a precision 0.1% or better by measuring the Bhabha rate in the polar-angle region from 30-90 mrad, and at smaller polar angles of 5-30 mrad the rate or energy deposition of beamstrahlung e+e- pairs can be measured for a fast luminosity diagnostic. The spatial distributions of pairs in this region can be used to determine beam collision parameters such as transverse sizes and bunch lengths. The MDI design must be compatible with stringent requirements on stability of the elements, in particular the final doublets. For the particular case when the linear collider will have a single interaction region but will be equipped with two experimental detectors in the so called push-pull configuration, the MDI region needs to be designed to be compatible with rapid disconnection and exchange of the detectors.

**Extraction and beam dump** The last element of the FFS is the primary beam dump. In the next generation of linear colliders [7, 12], typical average beam powers are many MW. After the beam-beam collision, a significant fraction of this power is in the form of photons. It is very important that the dump line transmits most particles to the dump. This also reduces background sources in the detector. Complete transmission is difficult to attain because of the enlarged beam emittance and energy tail due to beam-beam disruption. Thus one requirement of the extraction line is to have large apertures and a large energy bandwidth while capturing the disrupted beam.

Another consideration is that the dump line must include the diagnostics necessary to monitor and tune the beam collisions. The primary diagnostics needed are: (i) the beam position and angle after IP which are needed to interpret the beam-beam deflection scans to determine the IP beam sizes, (ii) the beam energy and energy spectrum, (iii) the beam polarization, and (iv) the

beam disruption. The last three items usually require separating the particle beam from the photon beam.

The dump must be able to withstand the average beam power as well as instantaneous energy deposition from a single bunch or bunch train. The heating due to the very high energy density in a single pulse generates a shock wave that can exceed the material stress limit and thus determines the minimum beam size at the dump. The main approach considered is to use a liquid such as water for the primary absorbing material similar to the beam dump developed for the high current 25 GeV beam at SLAC [43, 44].

## References

- [1] T.O. Raubenheimer, F. Zimmermann, Operation of Final-Focus Systems in Linear Colliders, in Handbook of Accelerator Physics and Engineering, 1st edition (1999); RMP 72 (2000) 95
- [2] V. Balakin et al, PRL 74 (1995) 2479
- [3] M. S. Zisman, LBL-34556 (1993)
- [4] E. Kikutani et al, KEK Preprint 2001-157
- [5] P. Raimondi, A. Seryi, PRL 86 (2001) 3779
- [6] P. Bambade et al (ATF Collaboration), PRST-AB 13, 042801 (2010)
- [7] ILC Reference Design Report, v.3, SLAC-R-857 (2007)
- [8] V.E. Balakin, 77th ICFA Workshop on Beam Dynamics, May 1991, Los Angeles
- [9] P. Chen, K. Yokoya, SLAC-PUB-4653 (1988)
- [10] R. Palmer, Snowmass DPF Summer Study 1988 (1988) p.613
- [11] R. Brinkmann, 5th Int. Work. Next Generation Linear Colliders, SLAC-436 (1993)
- [12] "A 3 TeV e+e- Linear Collider Based on CLIC Technology," CERN 2000-008
- [13] Y. Nosochkov, A. Seryi, PRST-AB 8, 021001 (2005)
- [14] B. Parker, A. Seryi, PRST-AB 8, 041001 (2005)
- [15] A. Seryi, T. Maruyama, B. Parker, Nanobeam 2005, SLAC-PUB-11662 (2006)
- [16] G. Rumolo, A. Latina, D. Schulte, EPAC 06
- [17] K.L.F. Bane, A. Seryi, PAC 07
- [18] G. Burt et al, EPAC 06
- [19] K. Yokoya, CERN SL/90-88 (1988)
- [20] H. Bethe, J. Ashkin, Experimental Nuclear Physics, Vol. 1, Wiley (1953)
- [21] I. Reichel et al, ICAP 98, eConf C980914 (1998) 222
- [22] F. Poirier et al, EPAC'08; W. Bialowons et al, EPAC06; A. Seryi, L. Hendrickson, G. White, Nanobeam 2005, SLAC-PUB-11661

## Sec.4.7: OPERATION OF CIRCULAR ACCELERATORS

- [23] J. Resta-López et al, JINST 5 P09007 (2010)
- [24] B.I. Grishanov et al, KEK-REPORT-2005-9 (2005)
- [25] B. Dunham, O. Napol, DAPNIA/SEA 94-06
- [26] A. Seryi, P. Raimondi, SLAC-PUB-9895, (2003)
- [27] H. Grote, F. C. Iselin, CERN SL-90-13-AP-REV.2 (1991)
- [28] D.C. Carey, FERMILAB-NAL-064 (1971)
- [29] R. Tomas, PRST-AB 13, 014801 (2010), and references therein
- [30] R. Tomas, PRST-AB 9, 081001 (2006)
- [31] R. Tomas, CERN-AB-Note-2006-017 (2006)
- [32] H. Grote, F. Schmidt, CERN-AB-2003-024
- [33] E. Forest, F. Schmidt, E. McIntosh, KEK Report 2002-3
- [34] D. Schulte, CERN PS-2000-028-AE (2000), <https://savannah.cern.ch/projects/placet/>
- [35] R. Servranckx et al, SLAC-R-285 (1990)
- [36] P. Tenenbaum, PAC 05, 4197
- [37] D. Schulte, TESLA-97-08 (1996)
- [38] Y. Nosochkov et al, SLAC-PUB-9255 (2005); G. White et al, EPAC 08
- [39] J. Smith et al, PAC 05
- [40] P. Tenenbaum et al, PAC 01
- [41] B.I. Grishanov et al, SLAC-R-771 (2005)
- [42] P. Bambade et al., PRST-AB 13, 04280 (2010)
- [43] D.R. Walz, IEEE Tr.Nucl.Sci. 3 NS-14 (1967) 923
- [44] Zeroth-Order-Design Report for the Next Linear Collider, LBNL-PUB-5424, SLAC-474 (1996)

## 4.7 OPERATION OF CIRCULAR ACCELERATORS

### 4.7.1 Error Sources and Effects [1, 2, 3]

*D. Rice, Cornell U.*

**Dipole errors** Transverse orbit displacement at position  $s$  due to deflections  $\theta_i$  is

$$\Delta x(s) = \frac{\sqrt{\beta(s)}}{2 \sin(\pi\nu)} \sum_{i=1}^N \theta_i \sqrt{\beta(s_i)} \times \cos(|\phi(s) - \phi(s_i)| - \pi\nu) \quad (1)$$

Resulting change in orbit circumference is

$$\Delta L = \sum_{i=1}^N \theta_i D(s_i) \quad (2)$$

with  $D(s)$  the dispersion function.

The displacement and slope of the closed orbit due to a single deflection, at the location of the

deflection source is

$$\Delta x(s_i) = \frac{\theta_i}{2} \beta(s_i) \cot(\pi\nu) \quad (3)$$

$$\Delta x'(s_i^+) = \frac{\theta_i}{2} [1 - \alpha(s_i) \cot(\pi\nu)] \quad (4)$$

where  $\Delta x'(s_i^+)$  is the slope of the orbit immediately after the kick. The maximum orbit excursions

$$\Delta \hat{x}(s) = \frac{\sqrt{\beta(s_i)\beta(s)}}{2 \sin(\pi\nu)} \theta_i \quad (5)$$

occur at  $\Delta\phi = \pm\pi(\nu-m)$ ,  $m = 1, 2, 3, \dots < \nu$ .

The probable displacement of the closed orbit at location  $s$  from a random distribution of  $N$  deflections is

$$\langle \Delta x(s) \rangle_{\text{rms}} = \sqrt{\beta(s)} \frac{\sqrt{N \langle \beta(s_i) \rangle}}{2\sqrt{2} \sin(\pi\nu)} \langle \theta_i \rangle_{\text{rms}} \quad (6)$$

If the deflections are caused by displaced quadrupoles,

$$\theta_i = (Kl)_i \delta x_i \quad (7)$$

where  $1/(Kl)_i$  is the focal length [ $K = (c/E_0)(\partial B_y/\partial x)$ ] and  $\delta x_i$  is the quad displacement.  $K > 0$  for a defocusing quad.

**Quadrupole errors** The perturbation to  $\beta$ -function from  $N$  quadrupole errors at positions  $s_i$  ( $i = 1, \dots, N$ ) is

$$\frac{\Delta\beta(s)}{\beta(s)} = \frac{1}{2 \sin(2\pi\nu)} \sum_{i=1}^N (\Delta Kl)_i \beta(s_i) \times \cos(2|\phi(s) - \phi(s_i)| - 2\pi\nu) \quad (8)$$

The betatron tune change  $\Delta\nu = \nu - \nu_0$  is

$$\cos(2\pi\nu) - \cos(2\pi\nu_0) \quad (9)$$

$$= \frac{1}{2} \sin(2\pi\nu_0) \int_s^{s+C} \beta(s') \Delta K(s') ds'$$

For a single gradient error and small  $\Delta\nu$ ,

$$\Delta\nu = -\frac{\beta(s_i)}{4\pi} (\Delta Kl)_i \quad (10)$$

### Vertical dispersion errors [4]

$$D_y(s) = \frac{\sqrt{\beta_y(s)}}{2 \sin(\pi\nu_y)} \int_s^{s+C} F(s') \sqrt{\beta_y(s')} \times \cos[\phi(s') - \phi(s) - \pi\nu] ds' \quad (11)$$

$$F(s) = (K + SD_x)y_c - K_{sq}D_x - G_y$$

where  $S, K_{sq}, G_y$  are sextupole ( $\text{m}^{-2}$ ), skew quad ( $\text{m}^{-1}$ ) and dipole ( $\text{m}^{-1}$ ) strengths, and  $y_c$  is the closed orbit displacement.

The vertical emittance from vertical dispersion is

$$\epsilon_y(\text{disp}) = \frac{2J_\epsilon \langle D_y^2 \rangle}{J_y \beta_y} \sigma_\delta^2 \quad (12)$$

with  $J_{\epsilon,y}$  the damping partition numbers,  $\sigma_\delta$  the rms beam relative energy spread.

There are five terms in the expression for  $\langle D_y^2 \rangle / \beta_y$  [4]. The first three of which arise from (1) quad rotations  $\theta$ , (2) sextupole misalignments  $y_m$ , and (3) vertical dipole kicks, respectively:

$$(1) \frac{1}{2 \sin^2 \pi \nu_y} \sum_{\text{quads}} (Kl)^2 \theta^2 \beta_y D_x^2$$

$$(2) \frac{1}{8 \sin^2 \pi \nu_y} \sum_{\text{sexts}} (Sl)^2 y_m^2 \beta_y D_x^2$$

$$(3) \frac{\langle y_c^2 \rangle}{\beta_y} \quad \text{with corrected closed orbit}$$

The remaining terms include the closed orbit correlation function [4] and decrease with the number of correctors.

**Eddy current multipoles** Eddy currents induced in a changing magnetic field generate multipoles which sometimes need to be considered. For example, in a rippling dipole field  $B(t) = B \sin \omega t$ , a round beam pipe shields the field and the rippling field seen in the beam pipe by the beam becomes [5]

$$B(t) = \frac{B}{\sqrt{1 + \omega^2 \tau^2}} \sin[\omega t - \tan^{-1}(\omega \tau)] \quad (13)$$

with  $\tau[\text{s}] = \mu_0 \sigma_c b t / 2 = 2\pi \times 10^{-13} b[\text{mm}] t[\text{mm}] / \rho_r [\Omega \text{m}]$ , where  $b$  and  $t$  are the radius and thickness of the beam pipe, and  $\sigma_c$  its conductivity. Similarly, in a rippling quad  $K(t) = K \sin \omega t$ , the quad field seen by the beam is

$$K(t) = \frac{K}{\sqrt{1 + \omega^2 \tau^2 / 4}} \sin \left[ \omega t - \tan^{-1} \frac{\omega \tau}{2} \right] \quad (14)$$

In case the beam pipe is not round, a changing multipole can also generate higher multipole fields. The sextupole field generated in a changing dipole, e.g., is an operational consideration during ramping (particularly in fast cycling machines). If the beam pipe is a wide chamber with gap height  $g$  and thickness  $t$ , the sextupole field is [6]

$$\frac{\partial^2 B_y}{\partial x^2} = 2\mu_0 \sigma_c \frac{t}{g} \dot{B} \quad (15)$$

If the beam pipe is slightly elliptical with  $\epsilon = [(\text{width}) - (\text{height})]/(\text{height})$  ( $|\epsilon| \ll 1$ ), then [7]

$$\frac{\partial^2 B_y}{\partial x^2} = \mu_0 \sigma_c \frac{t}{4b} \dot{B} \epsilon \quad (16)$$

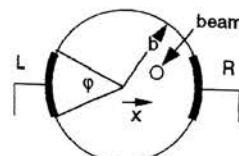
## References

- [1] E.D. Courant, H.S. Snyder, Ann. Phys. 3 (1958) 1
- [2] C. Bovet et al, CERN/MPS-SI/Int. DL/70/4 (1970)
- [3] H. Wiedemann, Particle Accelerator Physics, Vol.1, Springer-Verlag (1993)
- [4] T.O. Raubenheimer, R.D. Ruth, PAC 89, 1435
- [5] R.E. Shafer, Fermilab report TM-991 (1980)
- [6] D.A. Edwards, M.J. Syphers, An Introduction to the Physics of High Energy Accelerators, Wiley (1993) p.113
- [7] A. Chao, SSC-N-434 (1987)

### 4.7.2 Orbit and Lattice Function Measurements

D. Rice, Cornell U.

**Beam position measurement** Beam position is usually measured by coupling to the EM field of the beam. Insulated plates or buttons, striplines, and rf cavities are the most commonly used couplers (Sec.7.4.6). The difference between two button pickups is normalized to the sum as [1]:



$$\frac{R - L}{R + L} = \frac{\frac{x}{b} \left( 4 \sin \left( \frac{\phi}{2} \right) \right)}{\phi} \quad (1)$$

with variables defined in the figure.

Alternatively, the phase and amplitude of a transverse mode in an rf cavity are processed to yield a position. A comparison of advantages and resolutions among these couplers is given in Tab.1. The resolution quoted is without averaging or numeric processing and is primarily determined by intrinsic bandwidth and coupling strength. Electronics noise figure, processing algorithms, and mechanical stability of the pickup will ultimately determine usable resolution.

Processing electronics vary according to needs (bunch-by-bunch, multi-turn vs. high-resolution averaging) and pickup type. Direct sampling (better time resolution) and rf down conversion (better position resolution) methods are shown in Fig.1.

Cavity BPM electronics are conceptually simple amplifier/down conversion circuits. However,

## Sec.4.7: OPERATION OF CIRCULAR ACCELERATORS

Table 1: Beam Position Monitor pickups compared (electrode spacing is for reference cited).

| Pickup Type      | Resolution ( $\mu\text{m}$ ) | Electrode Spacing | Time Response (ns) | Cost / monitor |
|------------------|------------------------------|-------------------|--------------------|----------------|
| Button [2]       | 10                           | 50 mm             | $\approx 1$        | modest         |
| Stripline [3]    | .15                          | 35 mm             | $\approx 2L/c$     | medium         |
| RF Cavity [4, 5] | 0.5                          | —                 | $\approx 50$       | higher         |

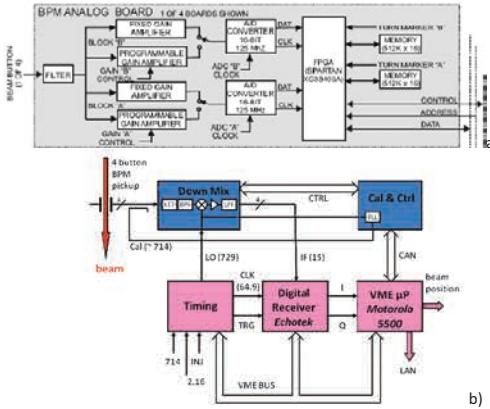


Figure 1: a) Direct sampling method [2], b) Frequency down conversion [6].

phase and amplitude stability throughout is critical to realize full resolution.

Appropriate signal processing methods such as averaging or Model Independent Analysis (MIA) using SVD (Sec.4.8) can give one or two orders of magnitude improvement in usable beam position measurement resolution as well as identify faulty detectors [6, 7]. Resolutions below 30 nm have been achieved using cavity BPM pickups and MIA/SVD.

**Dispersion Measurement** The dispersion function  $D(s)$  can be measured at each BPM by subtracting 2 orbits taken at different rf frequencies. Knowing the phase slip factor  $\eta$ ,

$$\begin{aligned} \Delta x(s) &= D(s) \frac{\Delta E}{E_0} \\ &= \frac{D(s)}{\eta} \frac{\Delta C}{C} = -\frac{D(s)}{\eta} \frac{\Delta f_{rf}}{f_{rf}}. \end{aligned} \quad (2)$$

A variation of this method is to drive the beam longitudinally at its synchrotron frequency and to process as in a coupling measurement (Sec.4.7.8), using the horizontal signal in a dispersive region as a reference [8, 9, 10]. In addition to being faster, this method is useful when aperture limitations restrict the swing in rf frequency. (Caution

must be exercised when  $Q_\beta - Q_s$  separation is not large [9].)

**Beam centering in quadrupoles** Modulating a quadrupole by a small amount  $\Delta K$  when the beam is displaced  $\Delta x$  from the magnetic center will cause a deflection  $\Delta\theta_Q$

$$\Delta\theta_Q = -\Delta x \Delta K l$$

where  $l$  is the magnet length.  $\Delta x$  may then be calculated from the measured change in the closed orbit. In practice, because of the change in optics, closed orbit bumps are successively applied until the orbit no longer changes as the quadrupole current is changed. This method is difficult to implement when the BPM pickups are located any significant distance from the quadrupoles. The change in beam closed-orbit coordinates at a quadrupole (thin lens approximation) resulting from a change in the quadrupole gradient is [11]:

$$\Delta x_{co|exit} = -x_{bq} \frac{\beta \Delta K \cot(\pi\nu)}{2 + \beta \Delta K \cot(\pi\nu)}$$

$$\Delta p_{co|exit} = -x_{bq} \frac{\Delta K [1 - \alpha \cot(\pi\nu)]}{2 + \beta \Delta K \cot(\pi\nu)}$$

where  $x_{bq}$  is the initial displacement of the closed orbit from the quad center,  $\beta$ ,  $\alpha$  are the usual twiss parameters at the quadrupole exit,  $\Delta K$  is the change in quadrupole strength, and  $\nu$  is the betatron tune. A transfer map may be constructed relating measured changes in closed orbit to the (unknown)  $x_{bq}$ 's and (known)  $\Delta K$ 's, from which the  $x_{bq}$ 's may be found. Adding betatron phase advance measurements to the orbit difference measurements improves knowledge of the optics beta functions, enhancing the resolution of this method [12].

**Beam centering in sextupoles** A sextupole of strength  $S$  is modulated by  $\Delta S$ . The changes in betatron tunes are proportional to the horizontal displacement of the beam,

$$\Delta\nu_{x,y} = \mp \frac{\beta_{x,y}(s)}{4\pi} \Delta x \Delta S l$$

and with vertical displacement the minimum split between the transverse normal modes is given by

$$\delta\nu_{min} = \frac{1}{2\pi} \Delta y (\Delta S l) \sqrt{\beta_x \beta_y}$$

Several accelerators have skew-quad windings on sextupoles, permitting the magnetic center of the skew quad (presumed close to the sextupole center) to be determined by techniques similar to those for quadrupoles [13] with the appropriate cross terms:

$$\theta_{x,y} = -K_S \Delta(y, x)$$

where  $\theta_{x,y}$  is the kick in  $x$  or  $y$  plane from a displacement in the  $y$  or  $x$  plane respectively and  $K_S$  the sextupole strength, to be varied. 80  $\mu\text{m}$  typical error is estimated in reference [13].

**Beta function measurement** The  $\beta$  function may be directly measured at each quadrupole by modulating the current in that quadrupole and measuring the change in betatron tune using

$$\Delta\nu = -\frac{\beta(s)}{4\pi} \Delta K(s) l$$

The accuracy of this measurement is degraded by hysteresis in the magnet field and potentially by coupling resonances. It is particularly unsatisfactory at locations where  $\beta(s)$  is small.

The beam may be “kicked” and its position measured on several successive turns, in which case the envelope of the free betatron oscillation follows  $\sqrt{\beta_{x,y}}$ . Alternatively measurements of several non-degenerate closed orbit errors may be fitted with  $\beta$  and the corresponding phase  $\varphi_{x,y}(s)$  as free parameters (Sec.4.7.5).

A direct measurement of betatron phase advance may be accomplished by shaking the beam at its betatron frequency and measuring the signal phase  $\theta_S$  at each BPM [12]. The betatron phase advance between a point  $s_i$  and a reference point  $s_0$  is:

$$\varphi(s_i) - \varphi(s_0) = \int_{s_0}^{s_i} \frac{1}{\beta(s)} ds$$

This method is particularly useful where  $\beta$  is small or where beam aperture is limited, and can be carried out with minimal perturbation to machine operation.

In analogue systems the measured phase  $\theta_S(i)$  contains contributions from the beam time of flight, as well as cable delays. Here one must make a baseline measurement,  $\theta_{rev}(i)$ , at the revolution frequency (without excitation). The betatron phase is then:

$$\begin{aligned} \varphi(i) &= \theta_s(i) + 2\pi m(i) + \\ &[\theta_{rev}(i) - 2\pi p(i)] \frac{\omega_{sb}}{\omega_0} + \theta \end{aligned}$$

where  $\theta$  is an overall arbitrary constant,  $p(i)$  is an integer chosen so that  $2\pi p(i) - \theta_{rev}(i)$  gives the correct phase shift, adjusting for time of flight and cable delays within a factor of  $\pi$ , and  $m(i)$  is an integer chosen so that  $\phi$  (measured)  $- \phi$  (theory)  $< \pi$ .  $\omega_{sb}$  is the measured betatron frequency in the relevant plane. If  $\omega_{sb} < 0$  (lower sideband) is measured, the sign of  $\theta_s(i)$  must be changed. With digital sampling of the BPM signal the time of flight correction is intrinsic in the sampling process.

All buttons should be measured and averaged with proper sign since local coupling will affect the betatron phase measurement at a single button. Typical achievable resolution is  $1^\circ$  for a single measurement and  $\sim 0.2^\circ$  with averaging [14].

**Tune measurement** The transverse and longitudinal motions of the beam in a periodic accelerator are sampled by a monitor at a single azimuth. For a bunched beam this sampling produces a series of spectral lines at frequencies  $f_i = |if_0 + f_{\beta,s}|$  ( $i$ =integer; the negative frequencies become the lower sidebands). The useful frequency spectrum is limited by the finite beam bunch length or pickup electrode characteristics.

A simple but highly sensitive detection scheme is shown in Fig.2 [15]. The fast signals from a pair of pickups are peak detected then differentially amplified. The time constant  $R_f C_f$  is matched to the required bandwidth. Measurement resolution can approach  $10^{-5}$  in tune fraction (sideband frequency) with this method.

Hadron storage rings require special consideration [16] to minimize emittance growth. Proton and ion storage rings may use Schottky spectra for both bunched and unbunched beam diagnostics. (Sec.7.4.1.7)

Once a high resolution tune spectrum is in hand, a wide variety of measurements follow, for example: chromaticity ( $d\nu/(dp/p_0)$ ) (Sec.2.1.1); beta functions (see above); optics nonlinearity (Sec.2.3.2); impedances (Sects.7.5.2 and 7.5.3).

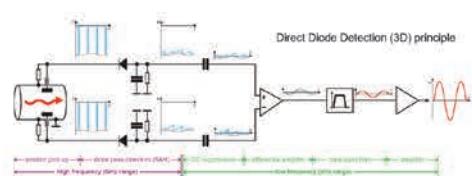


Figure 2: Direct Diode Detection for tune measurement [13].

### References

- [1] R. Shafer, AIP Proc. 249 (1992) 601
- [2] M. Palmer et al, IPAC 10, 1191
- [3] X.J. Wang, I Ben-Zvi, Z. Segalov, EPAC 96, 1576
- [4] R. Lill et al, PAC 07, 4366
- [5] S.T. Boogert et al, IPAC 10, 1140
- [6] M. Wendt et al, DIPAC 09
- [7] J. Irwin et al, PRL 82 (1999) 8
- [8] J. Borer et al, EPAC 92, 1082
- [9] F. Ruggiero, CERN SL-91-38-AP
- [10] J. Shanks et al, PAC 09, WE6PFP104
- [11] A. Wolski, F. Zimmermann, CERN-AB-2004-010, LBNL Report - 54360, ATF-03-08 (2004)
- [12] J. Shanks et al., IPAC 10, 4620
- [13] N. Terunuma et al, ICFA Beam Dynamics Panel Newsletter 50 (2009) 36
- [14] D. Sagan et al, EPAC 96, 1335
- [15] M. Gasior et al, BIW 10, TUCNB01
- [16] J. Klute et al, DIPAC 97 Frascati, 230

### 4.7.3 Orbit Correction

*J. Wenninger, CERN*

#### 4.7.3.1 Global orbit correction

Consider a storage ring with  $M$  beam position monitors (BPM) and  $N$  correctors. Orbit displacements  $\vec{d}$  ( $M$ -component vector) arising from corrector kick angles  $\vec{\theta}$  ( $N$ -component vector) are determined by the  $M \times N$  linear response matrix  $\mathbf{A}$ ,

$$\mathbf{A}\vec{\theta} = \vec{d} \quad (1)$$

$$A_{mn} = \frac{\sqrt{\beta_m \beta_n}}{2 \sin \pi \nu} \cos(|\phi_m - \phi_n| - \pi \nu)$$

The elements of  $\mathbf{A}$  may be obtained from the machine model or be determined experimentally by measuring deviation at each BPM resulting from exciting each corrector individually.

The task of the orbit correction is to find a set of corrector kicks  $\vec{\theta}$  that satisfy the following relation,

$$\vec{d} + \mathbf{A}\vec{\theta} = 0. \quad (2)$$

In general the number of BPMs ( $M$ ) and the number of correctors ( $N$ ) are not identical and Eq.(2) is either over- ( $M > N$ ) or under-constrained ( $M < N$ ). In the former and most frequent case, Eq.(2) can not be solved exactly. Instead, an approximate solution must be found, and commonly used least square algorithms minimize the quadratic residual

$$S = \|\vec{d} + \mathbf{A}\vec{\theta}\|^2. \quad (3)$$

#### 4.7.3.2 SVD algorithm

When  $M \geq N$ , the *Singular Value Decomposition* (SVD) [1] of matrix  $\mathbf{A}$  has the form  $\mathbf{A} = \mathbf{U}\mathbf{W}\mathbf{V}^t$ ,

$$\mathbf{A} = \begin{bmatrix} u_1^{(1)} & u_1^{(2)} & \cdots & u_1^{(N)} \\ u_2^{(1)} & u_2^{(2)} & \cdots & u_2^{(N)} \\ \vdots & \vdots & & \vdots \\ u_M^{(1)} & u_M^{(2)} & \cdots & u_M^{(N)} \end{bmatrix} \begin{bmatrix} w_1 & & & \\ & w_2 & & \\ & & \ddots & \\ & & & w_N \end{bmatrix} \times \begin{bmatrix} v_1^{(1)} & v_2^{(1)} & \cdots & v_N^{(1)} \\ v_1^{(2)} & v_2^{(2)} & \cdots & v_N^{(2)} \\ \vdots & \vdots & & \vdots \\ v_1^{(N)} & v_2^{(N)} & \cdots & v_N^{(N)} \end{bmatrix} \quad (4)$$

$\mathbf{U}$  is the  $M \times N$  matrix whose column vectors  $\vec{u}^{(\alpha)}$ , ( $\alpha = 1, \dots, N$ ) form an orthonormal set,  $\mathbf{U}^t \mathbf{U} = \mathbf{I}$ .  $\mathbf{W}$  is  $N \times N$  diagonal matrix with non-negative elements.  $\mathbf{V}^t$  is the transpose of the  $N \times N$  matrix  $\mathbf{V}$ , whose column vectors  $\vec{v}^{(\alpha)}$ , ( $\alpha = 1, \dots, N$ ) form an orthonormal set,  $\mathbf{V}^t \mathbf{V} = \mathbf{V}\mathbf{V}^t = \mathbf{I}$ .

From Eq.(4), it follows that ( $\alpha = 1, \dots, N$ ),

$$\mathbf{A}\vec{v}^{(\alpha)} = w_\alpha \vec{u}^{(\alpha)}, \quad \mathbf{A}^t \vec{u}^{(\alpha)} = w_\alpha \vec{v}^{(\alpha)} \quad (5)$$

and

$$\mathbf{A}\mathbf{A}^t \vec{u}^{(\alpha)} = w_\alpha^2 \vec{u}^{(\alpha)}, \quad \mathbf{A}^t \mathbf{A} \vec{v}^{(\alpha)} = w_\alpha^2 \vec{v}^{(\alpha)} \quad (6)$$

When none of the diagonal elements  $w_\alpha$  vanish, the solution of Eq.(2) is  $\vec{\theta} = -\mathbf{V}\mathbf{W}^{-1}\mathbf{U}^t \vec{d}$ .  $\vec{d}$  may be expanded in terms of eigenvectors  $\vec{u}^{(\alpha)}$  [2],

$$\vec{d} = \sum_{\alpha=1}^N C_\alpha \vec{u}^{(\alpha)} + \vec{d}_0 \quad (7)$$

where  $C_\alpha = \vec{d} \cdot \vec{u}^{(\alpha)}$ , while  $\vec{d}_0$  corresponds to the uncorrectable part of the orbit. The corrector strength required for correction is

$$\vec{\theta} = -\sum_{\alpha=1}^N \frac{C_\alpha}{w_\alpha} \vec{v}^{(\alpha)} \quad (8)$$

If a given  $w_\alpha = 0$  indicating that the matrix is singular, one discards the corresponding term from Eq.(8). An example for an eigenvalue spectrum is given in Fig.1 for LEP.

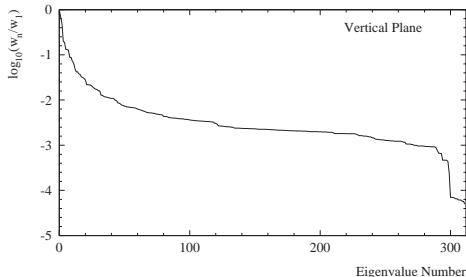


Figure 1: Vertical orbit eigenvalue spectrum for LEP. The last 4 eigenvalues correspond to singular solutions in the low-beta sections around the interaction points.

In practice one may want to limit the number of eigenvalues used for the correction to control the r.m.s. strength of the orbit correctors or to avoid small eigenvalues that are very sensitive to the accuracy of the model.

The SVD algorithm is ideally suited for feedback application since the correction can be cast in the simple form of a matrix multiplication once the SVD decomposition has been performed. This provides a fast and reliable correction procedure for a realtime feedback.

#### 4.7.3.3 MICADO algorithm

MICADO [3] is a least square correction algorithm based on Householder transformations. MICADO performs an iterative search for the most effective corrector and is, together with SVD, one of the most common orbit correction algorithms. For a non-singular matrix, a MICADO correction with all  $N$  correctors and an SVD correction with all  $N$  eigenvectors yield identical solutions. For corrections with a limited number of correctors or eigenvectors, and for singular matrices, the two algorithms converge differently.

A major difference between SVD and MICADO is the corrector strength distribution, MICADO using fewer but also much stronger kicks. The corrector strength r.m.s. can be easily controlled with SVD over the number of eigenvalues that are included in a correction. A correction of a small number of localized kicks is very effectively handled by MICADO, particularly when the response matrix is accurate, in which case MICADO can be used to identify the sources of the kicks. On the other hand, corrections based

on few eigenvectors with the largest eigenvalues are similar to corrections of the main harmonics. Such a scheme spreads out the correction of a few kicks over the whole machine which can be an asset when the strength of correctors is limited. To compensate an isolated kick locally, a large number of eigenvectors must be included in the correction such that the linear combination forming  $\vec{\theta}_c$  converges to a single nonzero corrector.

Singularities of the response matrix, associated to very small eigenvalues, are handled more easily with SVD, since it is sufficient to avoid using the corresponding eigenvectors in the corrections procedure. For the MICADO algorithm, it is necessary to regularize matrix  $\mathbf{A}$  by removing redundant correctors.

#### 4.7.3.4 Local orbit bumps

A local bump may be built from three correctors with deflections  $\theta_{1,2,3}$  at locations 1, 2, 3. The deflections may be expressed in terms of the lattice parameters,

$$\begin{aligned} \frac{\theta_2}{\theta_1} &= -\sqrt{\frac{\beta_1}{\beta_2}} \frac{\sin(\phi_3 - \phi_1)}{\sin(\phi_3 - \phi_2)} \\ \frac{\theta_3}{\theta_1} &= -\sqrt{\frac{\beta_1}{\beta_3}} \frac{\sin(\phi_2 - \phi_1)}{\sin(\phi_2 - \phi_3)} \end{aligned} \quad (9)$$

At a target point  $t$  between 1 and 2, the position and angular displacements are

$$\begin{aligned} d_t &= \theta_1 \sqrt{\beta_1 \beta_t} \sin(\phi_t - \phi_1) \\ d'_t &= \theta_1 \sqrt{\frac{\beta_1}{\beta_t}} [\cos(\phi_t - \phi_1) \\ &\quad - \alpha_t \sin(\phi_t - \phi_1)] \end{aligned} \quad (10)$$

At a point  $t$  between 2 and 3,

$$\begin{aligned} d_t &= \theta_3 \sqrt{\beta_3 \beta_t} \sin(\phi_3 - \phi_t) \\ d'_t &= -\theta_3 \sqrt{\frac{\beta_3}{\beta_t}} [\cos(\phi_3 - \phi_t) \\ &\quad + \alpha_t \sin(\phi_3 - \phi_t)] \end{aligned} \quad (11)$$

To control both position  $d_t$  and angle  $d'_t$  at the source point, a four-magnet local bump is required. The four-magnet local bump with corrector locations 1, 2, 3, 4, where the source point  $t$  is located between correctors 2 and 3 is given in

## Sec.4.7: OPERATION OF CIRCULAR ACCELERATORS

terms of optics functions by:

$$\begin{aligned}\theta_1 &= \frac{d_t(\cos(\phi_t - \phi_2) - \alpha_t \sin(\phi_t - \phi_2))}{\sqrt{\beta_t \beta_1} \sin(\phi_2 - \phi_1)} \\ &\quad - \frac{d'_t \sqrt{\beta_t / \beta_1} \sin(\phi_t - \phi_2)}{\sin(\phi_2 - \phi_1)} \\ \theta_2 &= \frac{-d_t(\cos(\phi_t - \phi_1) - \alpha_t \sin(\phi_t - \phi_1))}{\sqrt{\beta_t \beta_2} \sin(\phi_2 - \phi_1)} \\ &\quad + \frac{d'_t \sqrt{\beta_t / \beta_2} \sin(\phi_t - \phi_1)}{\sin(\phi_2 - \phi_1)} \\ \theta_3 &= \frac{-d_t(\cos(\phi_4 - \phi_t) - \alpha_t \sin(\phi_t - \phi_4))}{\sqrt{\beta_t \beta_3} \sin(\phi_4 - \phi_3)} \\ &\quad + \frac{d'_t \sqrt{\beta_t / \beta_3} \sin(\phi_t - \phi_4)}{\sin(\phi_4 - \phi_3)} \\ \theta_4 &= \frac{d_t(\cos(\phi_3 - \phi_t) - \alpha_t \sin(\phi_t - \phi_3))}{\sqrt{\beta_t \beta_4} \sin(\phi_4 - \phi_3)} \\ &\quad - \frac{d'_t \sqrt{\beta_t / \beta_4} \sin(\phi_t - \phi_2)}{\sin(\phi_4 - \phi_3)}\end{aligned}\quad (12)$$

## References

- [1] W.H. Press et al, Numerical Recipes, Cambridge U. Press (1988) p.52
- [2] A. Friedman, E. Bozoki, NIM A344 (1994) 269
- [3] B. Autin, Y. Marti, CERN report ISR MA/73-17 (1973)

### 4.7.4 Measurement and Diagnosis of Coupling and Solenoid Compensation

*D. Rubin, Cornell U.*

In  $e^+e^-$  storage ring colliders, luminosity is inversely proportional to the vertical dimension of the ribbonlike beams. The most important source of vertical beam size is from its coupling to horizontal and longitudinal motion.

#### 4.7.4.1 Sources of transverse coupling

**Tilted quadrupoles** The  $4 \times 4$  transfer matrix for a horizontally focusing quad with strength  $k$  and length  $l$  can be written as (Sec.2.2.1)

$$\mathbf{M}_{\text{quad}} = \begin{bmatrix} \mathbf{K}_f & 0 \\ 0 & \mathbf{K}_d \end{bmatrix} \quad (1)$$

where  $\mathbf{K}_{f,d}$  are appropriate  $2 \times 2$  matrices. The matrix for a quad rotated about an angle  $\theta$  is

$$\mathbf{Q}_{\text{rot}} = \mathbf{R}^{-1}(\theta) \mathbf{M}_{\text{quad}} \mathbf{R}(\theta) \quad (2)$$

where

$$\mathbf{R}(\theta) = \begin{bmatrix} \mathbf{I} \cos \theta & \mathbf{I} \sin \theta \\ -\mathbf{I} \sin \theta & \mathbf{I} \cos \theta \end{bmatrix} \quad (3)$$

**Skew quadrupoles** A skew quad is a quad rotated by  $45^\circ$ ,

$$\mathbf{Q}_{\text{skew}} = \frac{1}{2} \begin{bmatrix} \mathbf{K}_f + \mathbf{K}_d & -\mathbf{K}_f + \mathbf{K}_d \\ -\mathbf{K}_f + \mathbf{K}_d & \mathbf{K}_f + \mathbf{K}_d \end{bmatrix} \quad (4)$$

For a thin skew quad,  $l \rightarrow 0$  and  $\sqrt{k} \sin(\sqrt{k}l) \rightarrow \frac{1}{f}$ ,

$$\mathbf{Q}_{\text{thin}} = \begin{bmatrix} \mathbf{I} & \mathbf{K}_t \\ \mathbf{K}_t & \mathbf{I} \end{bmatrix}, \quad K_t = \begin{bmatrix} 0 & 0 \\ \frac{1}{f} & 0 \end{bmatrix} \quad (5)$$

#### 4.7.4.2 Solenoids

**Longitudinal fields** The transfer matrix for the motion through a longitudinal field is

$$\mathbf{M}_{\text{long}} = \begin{bmatrix} \mathbf{M}_1^s & \mathbf{M}_2^s \\ -\mathbf{M}_2^s & \mathbf{M}_1^s \end{bmatrix} \quad (6)$$

where  $k_s = \frac{e}{2pc} B_z$  and

$$\mathbf{M}_1^s = \begin{bmatrix} 1 & \frac{1}{k_s} \sin k_s z \\ 0 & \cos k_s z \end{bmatrix} \quad (7)$$

$$\mathbf{M}_2^s = \begin{bmatrix} 0 & \frac{1}{k_s} (\cos k_s z - 1) \\ 0 & -\sin k_s z \end{bmatrix} \quad (8)$$

**Radial fringe** The fringe field of a solenoid is radial, of equal magnitude and opposite direction at each end. The elements of the transfer matrix for the radial fringe are

$$\begin{aligned}(x|x_0) &= \cos \chi \cosh \chi \\ (x|x'_0) &= \frac{1}{\sqrt{2k_r}} (\sin \chi \cosh \chi + \sinh \chi \cos \chi) \\ (x|y_0) &= \sin \chi \sinh \chi \\ (x|y'_0) &= \frac{1}{\sqrt{2k_r}} (\sin \chi \cosh \chi - \sinh \chi \cos \chi) \\ (y|x_0) &= -(x|y_0) \\ (y|x'_0) &= -(x|y'_0) \\ (y|y_0) &= \cos \chi \cosh \chi \\ (y|y'_0) &= \frac{1}{\sqrt{2k_r}} (\sin \chi \cosh \chi + \cos \chi \sinh \chi)\end{aligned}\quad (9)$$

where  $\chi = \sqrt{\frac{k}{2}} z$ ,  $k_r = \frac{1}{2a} \frac{e}{pc} B_z$ , and  $a$  is the length (along  $z$ ) of the pole tips.  $[(x'|x_0), (x'|x'_0)$ , etc. are obtained by differentiating  $(x|x_0), (x|x'_0)$ , etc. with respect to  $z$ .] In the limit of a thin radial fringe  $a \rightarrow z \rightarrow 0$  and  $k_r z = \frac{k_s z}{2a} \rightarrow \frac{k_s}{2}$ , the transfer matrix becomes

$$\mathbf{M}_{\text{fringe}} = \begin{bmatrix} \mathbf{I} & \mathbf{K}_s \\ -\mathbf{K}_s & \mathbf{I} \end{bmatrix}, \quad (10)$$

where

$$\mathbf{K}_s = \begin{bmatrix} 0 & 0 \\ \frac{k_s}{2} & 0 \end{bmatrix} \quad (11)$$

**Symplecticity of solenoid maps** Eqs.(6), (9) are not symplectic, but the solenoid matrix  $\mathbf{M}_{\text{sol}} = \mathbf{M}_{\text{fringe}} \mathbf{M}_{\text{long}} \mathbf{M}_{\text{fringe}}^{-1}$  is symplectic. See Eq.(21), Sec.2.2.1.

**Solenoid lens**  $\mathbf{M}_{\text{sol}}$  can also be written as a combination of rotations of an angle  $\theta = k_s l/4$  and a thick lens that focuses in both planes with focusing strength  $k = (k_s/2)^2$ , where  $k_s = \frac{e}{pc} B_z$  and  $l$  is the solenoid length,

$$\mathbf{M}_{\text{sol}} = \mathbf{R} \left( \frac{k_s}{4} l \right) \mathbf{F} \left( \frac{k^2}{4}, l \right) \mathbf{R} \left( \frac{k_s}{4} l \right) \quad (12)$$

$$\mathbf{F} = \begin{bmatrix} \mathbf{K}_f(k, l) & 0 \\ 0 & \mathbf{K}_f(k, l) \end{bmatrix}$$

$K_f$  is as in Eq.(1).

### Superimposed solenoid and quadrupole fields

The matrices  $\mathbf{M}_{QL}$  and  $\mathbf{M}_{QR}$  are given for superimposed quadrupole and longitudinal fields and for superimposed quadrupole and radial fields. The elements of  $\mathbf{M}_{QL}$  are

$$\begin{aligned} (x|x_0) &= -\frac{1}{fk_q} (g^+ \theta_-^2 \cos \theta_+ z - g^- \theta_+^2 \cosh \theta_- z) \\ (x|x'_0) &= \frac{1}{fk_q} (-g^+ \theta_- \sin \theta_+ z + g^- \theta_+ \sinh \theta_- z) \\ (x|y_0) &= \frac{k_q k_s}{f|k_q|} (-\theta_- \sin \theta_+ z + \theta_+ \sinh \theta_- z) \\ (x|y'_0) &= \frac{k_s}{f} (\cos \theta_+ z - \cosh \theta_- z) \\ (y|x_0) &= (x|y_0) \\ (y|x'_0) &= -(x|y'_0) \\ (y|y_0) &= \frac{1}{f} (g^- \cos \theta_+ z - g^+ \cosh \theta_- z) \\ (y|y'_0) &= \frac{1}{fk_q} (g^- \theta_+ \sin \theta_+ z + g^+ \theta_- \sinh \theta_- z) \end{aligned} \quad (13)$$

The focusing strength of the quadrupole is  $k_q$ ,  $k_s = \frac{e}{pc} B_z$ ,  $f = \sqrt{k_s^4 + 4k_q^2}$ ,  $\theta_\pm = |\sqrt{\frac{1}{2}(k_s^2 \pm f)}|$ , and  $g^\pm = k_q - \frac{1}{2}(k_s^2 \pm f)$ . Elements  $(x'|x_0)$ ,  $(x'|x'_0)$ , etc. are obtained by differentiating  $(x|x_0)$ ,  $(x|x'_0)$ , etc. with respect to  $z$ .

If  $k_q^2 > k_r^2$  (where  $k_r = k_s/(2a)$ ), elements of  $\mathbf{M}_{QR}$  are

$$\begin{aligned} (x|x_0) &= \frac{1}{h} (g^+ \cos \phi z - g^- \cosh \phi z) \\ (x|x'_0) &= \frac{2\phi}{h^2} (g^+ \sin \phi z - g^- \sinh \phi z) \\ (x|y_0) &= \frac{k_p}{h} (\cosh \phi z - \cos \phi z) \\ (x|y'_0) &= \frac{2k_r \phi}{h^2} (\sinh \phi z - \sin \phi z) \\ (y|x_0) &= -(x|y_0) \\ (y|x'_0) &= -(x|y'_0) \\ (y|y_0) &= \frac{1}{h} (g^+ \cosh \phi z - g^- \cos \phi z) \\ (y|y'_0) &= \frac{2\phi}{h^2} (g^+ \sinh \phi z - g^- \sin \phi z) \end{aligned} \quad (14)$$

If  $k_r^2 > k_q^2$ , then

$$\begin{aligned} (x|x_0) &= \frac{2}{h} (\phi^2 \cos \zeta \cosh \zeta + k_q \sin \zeta \sinh \zeta) \\ (x|x'_0) &= \frac{2\alpha}{h\phi^2} (-g^- \sin \zeta \cosh \zeta + g^+ \sinh \zeta \cos \zeta) \\ (x|y_0) &= \frac{2k_r}{h} (\sin \zeta \sinh \zeta) \\ (x|y'_0) &= \frac{-2k_r \alpha}{h\phi^2} (\cos \zeta \sinh \zeta - \sin \zeta \cosh \zeta) \\ (y|x_0) &= -(x|y_0) \\ (y|x'_0) &= -(x|y'_0) \\ (y|y_0) &= \frac{2}{h} (\phi^2 \cos \zeta \cosh \zeta - k_q \sin \zeta \sinh \zeta) \\ (y|y'_0) &= \frac{2\alpha}{h\phi^2} (g^+ \cosh \zeta \sin \zeta - g^- \sinh \zeta \cos \zeta) \end{aligned}$$

where  $\phi = |k_q^2 - k_r^2|^{\frac{1}{4}}$ ,  $h = 2\phi^2$ ,  $g^\pm = \pm\phi^2 - k_q$ ,  $\alpha = \phi/\sqrt{2}$  and  $\zeta = az$ . Again  $(x'|x_0)$ ,  $(x'|x'_0)$ , etc. are obtained by differentiating  $(x'|x_0)$ ,  $(x'|x'_0)$ , etc. with respect to  $z$ . Matrices  $\mathbf{M}_{QL}$  and  $\mathbf{M}_{QR}$  are not symplectic, but the combination  $\mathbf{M}_{QR}^{\text{thin}} \mathbf{M}_{QL} (\mathbf{M}_{QR}^{\text{thin}})^{-1}$  is symplectic.  $\mathbf{M}_{QR}^{\text{thin}}$  is  $\mathbf{M}_{QR}$  in the limit of zero length fringe. See Eq.(11).

**Sextupoles** When there is a vertical closed orbit  $y_0$  at a sextupole, the sextupole behaves as a skew quad with strength  $k = 2eB_0/(pcr_0^2)$ , where  $B_0$  and  $r_0$  are the field and the radius at the pole tip. The sensitivity of the luminosity in  $e^+e^-$  colliders to vertical orbit details is related to this property.

#### 4.7.4.3 Coupling matrix analysis

**Normal modes** The one-turn transfer matrix  $\mathbf{T}$  is decomposed into normal modes as [1, 2]

$$\mathbf{T} = \begin{bmatrix} \mathbf{M} & \mathbf{n} \\ \mathbf{m} & \mathbf{N} \end{bmatrix} = \mathbf{V} \mathbf{U} \mathbf{V}^{-1} \quad (15)$$

$$\text{where } \mathbf{U} = \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{B} \end{bmatrix}, \quad \mathbf{V} = \begin{bmatrix} \gamma \mathbf{I} & \mathbf{C} \\ -\mathbf{C}^\dagger & \gamma \mathbf{I} \end{bmatrix} \quad (16)$$

and  $\gamma^2 + \det \mathbf{C} = 1$ .  $\mathbf{A}$  and  $\mathbf{B}$  are the one-turn matrices for each of the two normal modes. Symplectic conjugate  $\mathbf{C}^\dagger = -\mathbf{S} \mathbf{C}^t \mathbf{S}$  (see also Sec.2.1.3). The laboratory phase space coordinates  $\mathbf{x} = (x, x', y, y')$  are related to the normal mode coordinates  $\mathbf{w} = (w, w', v, v')$  by  $\mathbf{x} = \mathbf{V}\mathbf{w}$ . The same relation holds for energy displacements so the normal mode dispersions may be calculated from  $(D_u, D'_u, D_v, D'_v)^t = \mathbf{V}^{-1} (D_x, D'_x, D_y, D'_y)^t$ . Given the normal mode Courant-Snyder parameters and dispersions the normal mode emittances can be calculated.

Phase space normalization of the normal mode vectors yields

$$\bar{\mathbf{w}} = \mathbf{G}\mathbf{w} = \begin{bmatrix} \mathbf{G}_A & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_B \end{bmatrix} \mathbf{w} \quad (17)$$

$$\mathbf{G}_{A,B} = \begin{bmatrix} \frac{1}{\sqrt{\beta_{A,B}}} & 0 \\ \frac{\sqrt{\alpha_{A,B}}}{\sqrt{\beta_{A,B}}} & \sqrt{\beta_{A,B}} \end{bmatrix}$$

We define

$$\bar{\mathbf{V}} \equiv \mathbf{G}\mathbf{V}\mathbf{G}^{-1} = \begin{bmatrix} \gamma\mathbf{I} & \bar{\mathbf{C}} \\ -\bar{\mathbf{C}}^\dagger & \gamma\mathbf{I} \end{bmatrix} \quad (18)$$

**Propagating the coupling matrix** Elements of the coupling matrix can be propagated through a lattice given the intervening transfer matrices. Consider the one-turn  $4 \times 4$  matrix at  $s_1$  with  $\mathbf{T}_1 = \mathbf{V}_1 \mathbf{U}_1 \mathbf{V}_1^{-1}$ . At another point  $s_2$ , we have

$\mathbf{T}_2 = \mathbf{V}_2 \mathbf{U}_2 \mathbf{V}_2^{-1}$ . Let  $\mathbf{T}_{12} = \begin{bmatrix} \mathbf{M}_{12} & \mathbf{m}_{12} \\ \mathbf{n}_{12} & \mathbf{N}_{12} \end{bmatrix}$  be the matrix that propagates from  $s_1$  to  $s_2$ , then

$\mathbf{U}_2 = \mathbf{W}\mathbf{U}_1\mathbf{W}^{-1}$  with  $\mathbf{W} = \mathbf{V}_2^{-1}\mathbf{T}_{12}\mathbf{V}_1$  (19) Since  $\mathbf{U}_{1,2}$  are block diagonal,  $\mathbf{W}$  is also block diagonal,  $\mathbf{W} = \begin{bmatrix} \mathbf{E}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{F}^{-1} \end{bmatrix}$ . (For very strong coupling,  $\mathbf{W}$  may be off block diagonal [3].) The coupling matrix at  $s_2$  can be written in terms of the elements of the coupling matrix at  $s_1$  by

$$\mathbf{C}_2 = (\mathbf{M}_{12}\mathbf{C}_1 + \gamma_1\mathbf{m}_{12})\mathbf{F}_{12}^{-1} \quad (20)$$

where

$$\gamma_2^2 = \det(\mathbf{n}_{12}\mathbf{C}_1 + \gamma_1\mathbf{N}_{12})$$

$$\mathbf{E}_{12} = (\gamma_1\mathbf{M}_{12} - \mathbf{m}_{12}\mathbf{C}_1^\dagger)/\gamma_2$$

$$\mathbf{F}_{12} = (\mathbf{n}_{12}\mathbf{C}_1 + \gamma_1\mathbf{N}_{12})/\gamma_2 \quad (21)$$

When there are no couplers between  $s_1$  and  $s_2$ ,  $\mathbf{n}$  and  $\mathbf{m}$  are zero and

$$\mathbf{C}_2 = \mathbf{M}_{12}\mathbf{C}_1\mathbf{F}_{12}^{-1} = \mathbf{M}_{12}\mathbf{C}_1\mathbf{N}_{12}^{-1} \quad (22)$$

Define  $\bar{\mathbf{M}}_{12} = \mathbf{G}_A^2 \mathbf{M}_{12} (\mathbf{G}_A^1)^{-1}$  and  $\bar{\mathbf{N}}_{12} = \mathbf{G}_B^2 \mathbf{N}_{12} (\mathbf{G}_B^1)^{-1}$  where  $\mathbf{G}_A^1$  is the normalization matrix for horizontal motion at  $s_1$ ,  $\mathbf{G}_A^2$  is the normalization matrix at  $s_2$ , etc. Then Eq.(22) gives

$$\mathbf{G}_A^2 \mathbf{C}_2 (\mathbf{G}_B^2)^{-1} = \bar{\mathbf{C}}_2 = \bar{\mathbf{M}} \bar{\mathbf{C}}_1 \bar{\mathbf{N}}^{-1} \quad (23)$$

The normalized transfer matrices  $\bar{\mathbf{M}}$  and  $\bar{\mathbf{N}}$  are simple rotations in the normalized phase space and  $\bar{\mathbf{C}}_2 = \mathbf{R}(\phi_A)\bar{\mathbf{C}}_1\mathbf{R}(\phi_B)^{-1}$ , and  $\mathbf{R}$  is a rotation matrix. Define complex coupling coefficients

$$a = \bar{c}_{11} - \bar{c}_{22} + i(\bar{c}_{12} + \bar{c}_{21})$$

$$b = \bar{c}_{11} + \bar{c}_{22} + i(\bar{c}_{21} - \bar{c}_{12}) \quad (24)$$

Then

$$a_2 = e^{-i(\phi_A + \phi_B)} a_1, \quad b_2 = e^{-i(\phi_A - \phi_B)} b_1 \quad (25)$$

where  $\phi_A + \phi_B$  is the sum of the normal mode phase advances between 1 and 2, and  $\phi_A - \phi_B$  is the difference.

**Coupling resonances** Suppose that a coupling error  $\delta a, \delta b$  is introduced at  $s_1$ . Then propagate  $a$  and  $b$  around the ring and back to  $s_1$  according to

$$e^{-i\Sigma} a = a + \delta a, \quad e^{-i\Delta} b = b + \delta b \quad (26)$$

$\Sigma = 2\pi(\nu_A + \nu_B)$  and  $\Delta = 2\pi(\nu_A - \nu_B)$ . The solution to Eq.(26) for  $a$  and  $b$  at  $s_1$  is

$$a = \frac{ie^{-i\frac{\Sigma}{2}}}{2\sin\frac{\Sigma}{2}} \delta a, \quad b = \frac{ie^{-i\frac{\Delta}{2}}}{2\sin\frac{\Delta}{2}} \delta b \quad (27)$$

Note that near the sum resonance the fast wave ( $a$ ) is magnified and near the difference resonance the slow wave ( $b$ ) tends to dominate.

#### 4.7.4.4 Measurement of coupling

**Tune split** If the full turn betatron phase advance for both horizontal and vertical motion can be tuned to the coupling resonance, the splitting of the normal mode tunes can be a useful measure of machine coupling. Eqs.(15, 16) give

$$\text{tr}(\mathbf{A} - \mathbf{B}) = \sqrt{\text{tr}(\mathbf{M} - \mathbf{N})^2 + 4 \det(\mathbf{m} + \mathbf{n}^\dagger)} \quad (28)$$

At the coupling resonance, where  $\nu_h = \nu_v$ ,  $\text{tr}(\mathbf{M} - \mathbf{N}) = 0$ , and Eq.(28) becomes

$$\cos 2\pi\nu_A - \cos 2\pi\nu_B = \sqrt{\det(\mathbf{m} + \mathbf{n}^\dagger)} \quad (29)$$

Thus the splitting of the normal mode tunes is proportional to  $\det(\mathbf{C})$ . If the full turn matrix is block diagonal, then  $\mathbf{m} = \mathbf{n} = \mathbf{0}$ , and the splitting of normal mode tunes can be reduced to zero. By adjusting coupling correctors, such as skew quads, to minimize the normal mode tune split near the sum resonance we are reducing  $a$  and minimizing the tune split at the difference resonance reduces  $b$ .

#### Tune split due to a thin skew quad or solenoid

Consider a machine tuned to the difference resonance that is described by the full turn block diagonal matrix at some point  $s$  by  $\mathbf{F} = \begin{bmatrix} \mathbf{M} & \mathbf{0} \\ \mathbf{0} & \mathbf{N} \end{bmatrix}$ .

We can write  $\mathbf{M} = \begin{bmatrix} \cos\mu_h & \beta_h \sin\mu_h \\ -\gamma \sin\mu_h & \cos\mu_h \end{bmatrix}$ , and similarly for  $\mathbf{N}$ . Here  $\mu_h = 2\pi\nu_h$ . At the difference resonance,  $\text{tr}(\mathbf{M} - \mathbf{N}) = 0$ . Then introduce a thin skew quad  $\mathbf{Q}_{\text{thin}}$  as in Eq.(5). The perturbed full turn matrix is

$$\mathbf{P} = \mathbf{F}\mathbf{Q} = \begin{bmatrix} \mathbf{M} & \mathbf{MK}_t \\ \mathbf{NK}_t & \mathbf{N} \end{bmatrix} \quad (30)$$

From Eq.(29) we have

$$\cos \mu_A - \cos \mu_B = \sqrt{\det[\mathbf{M}\mathbf{K}_t + (\mathbf{N}\mathbf{K}_t)^\dagger]} \quad (31)$$

Since  $\mu_h \sim \mu_v$ ,

$$\Delta\nu = \nu_A - \nu_B \sim \frac{1}{2\pi} \frac{\sqrt{\beta_h \beta_v}}{f} \quad (32)$$

To compute the tune split due to a thin solenoid, we substitute

$$\mathbf{M}_{\text{thin,sol}} = \begin{bmatrix} I & I \sin \theta \\ -I \sin \theta & I \end{bmatrix}$$

for  $\mathbf{Q}$  in Eq.(30) and then for the thin solenoid

$$\Delta\nu = \frac{1}{2\pi} \theta \frac{(\beta_h + \beta_v)}{\sqrt{\beta_h \beta_v}}$$

where  $\theta = k_s l$ , and  $k_s$  defined above.

**Coupling wave due to thin skew quads and solenoids** Eqs.(27, 28) and thin coupler matrices are combined to give a first order expression for the complex coupling parameters  $a$  and  $b$ . The coupling at  $j$  due to skew quads and/or solenoids at  $k$  is

$$a_j = \sum_k \rho_k e^{-i(\xi_{jk}^+ + \frac{\Sigma}{2})}, \quad b_j = \sum_k \chi_k e^{-i(\xi_{jk}^- + \frac{\Delta}{2})}$$

where

$$\rho_k = \frac{\sqrt{\beta_h^k \beta_v^k}}{\text{tr}(\mathbf{A} - \mathbf{B})\gamma} \left( \frac{2 \sin \frac{\Delta}{2}}{f_k} + \frac{2i \cos \frac{\Delta}{2}}{g_k} \right)$$

$$\chi_k = \frac{\sqrt{\beta_h^k \beta_v^k}}{\text{tr}(\mathbf{A} - \mathbf{B})\gamma} \left( \frac{2 \sin \frac{\Sigma}{2}}{f_k} + \frac{2i \cos \frac{\Sigma}{2}}{g_k} \right)$$

$$\xi_{jk}^\pm = \phi_h^j - \phi_h^k \pm (\phi_v^j - \phi_v^k)$$

$$\Sigma = \mu_h + \mu_v, \quad \Delta = \mu_h - \mu_v$$

In the limit of weak coupling  $\text{tr}(\mathbf{A} - \mathbf{B}) = \text{tr}(\mathbf{M} - \mathbf{N}) = 2(\cos \mu_h - \cos \mu_v)$ , and  $\gamma \sim 1$ .

**Relative amplitude and phase** When a normal mode is excited, the relative amplitude and phase of horizontal and vertical motion can be measured at a BPM [7]. Consider the motion in the laboratory coordinate system as a consequence of excitation of the  $A$  mode. If the initial normalized, normal mode vector is  $\mathbf{Gw}_0 = \bar{\mathbf{w}}_0 = (\sqrt{\epsilon_A}, 0, 0, 0)$ , after some number of turns when the normal mode  $A$  has propagated through some phase advance  $\phi$ , we have

$$\mathbf{x} = \mathbf{Vw}_n = \mathbf{VU}^n \mathbf{w}_0 = \mathbf{G}^{-1} \bar{\mathbf{V}} \bar{\mathbf{U}}^n \bar{\mathbf{w}}_0 \quad (33)$$

$$\begin{aligned} &\Rightarrow \begin{bmatrix} x \\ x' \\ y \\ y' \end{bmatrix} = \begin{bmatrix} \gamma \sqrt{\epsilon_A \beta_A} \cos \phi \\ x' \\ -\sqrt{\epsilon_A \beta_A} (\bar{c}_{22} \cos \phi + \bar{c}_{12} \sin \phi) \\ y' \end{bmatrix} \\ &\Rightarrow \begin{aligned} x &= \gamma \sqrt{\epsilon_A \beta_A} \cos \phi \\ y &= -\sqrt{\epsilon_A \beta_A} (\bar{c}_{22} \cos \phi + \bar{c}_{12} \sin \phi) \\ &= -\sqrt{\epsilon_A \beta_A} \sqrt{\bar{c}_{22}^2 + \bar{c}_{12}^2} \cos(\phi + \delta\phi) \end{aligned} \quad (34) \end{aligned}$$

where  $\delta\phi = \tan^{-1} \frac{\bar{c}_{12}}{\bar{c}_{22}}$ . The ratio of the horizontal and vertical amplitudes of the normal mode motion is

$$(y/x)_{\text{Amp}} = \frac{1}{\gamma} \sqrt{\frac{\beta_B}{\beta_A}} \sqrt{\bar{c}_{12}^2 + \bar{c}_{22}^2} \quad (35)$$

The ratios of the component of the vertical motion that is in phase and out of phase with the horizontal motion are

$$(y/x)_{\text{in phase}} = (1/\gamma) \sqrt{\beta_B / \beta_A \bar{c}_{22}} \quad (36)$$

$$(y/x)_{\text{out of phase}} = (1/\gamma) \sqrt{\beta_B / \beta_A \bar{c}_{12}} \quad (37)$$

The in-phase coupled motion corresponds to a tilt of the real space beam ellipse and the out of phase motion to an increase in the height of the ellipse.

Excitation of the orthogonal mode ( $\epsilon_B \neq 0$ ) yields

$$(x/y)_{\text{in phase}} = (1/\gamma) \sqrt{\beta_A / \beta_B} \bar{c}_{11}$$

$$(x/y)_{\text{out of phase}} = (1/\gamma) \sqrt{\beta_A / \beta_B} \bar{c}_{12}$$

**Measurement** The elements of the coupling matrix can be measured at BPM's. A magnetic shaker excites one or the other of the normal modes. The transfer function between shaker drive and horizontal and vertical motion at the BPM is measured. The relative phase of the vertical and horizontal motion  $\delta\phi = \bar{c}_{12}/\bar{c}_{22}$  and the relative amplitude is  $(y/x)_{\text{Amp}} = (1/\gamma) \sqrt{\beta_B / \beta_A} \sqrt{\bar{c}_{12}^2 + \bar{c}_{22}^2}$ . See also Sec.4.7.5.

The measurement of  $\bar{c}_{12}$  in CESR is reproducible at the 0.5% level. Coupling errors are typically reduced to  $< 1\%$ .  $\bar{c}_{22}$  is somewhat more difficult to measure since the vertical amplitude is typically  $\ll$  the horizontal amplitude.

#### 4.7.4.5 Solenoid compensation

Large detector solenoids are strong coupling elements. The compensation consists of coupling elements deployed to globally decouple the motion outside the compensation region and to preclude feedthrough of horizontal motion to beam height at the IP.

**Compensation with anti-solenoids** The coupling of the solenoid can be compensated by anti-solenoids of equal but opposite integrated strength. If a pair of half strength anti-solenoids are placed symmetrically about the main solenoid, then there is vanishing coupling at the mid-point of the main solenoid. The anti-solenoids may be displaced from the ends of the main solenoid by a field free drift. However, the solenoid matrix does not commute with that of a quadrupole. And anti-solenoids are ineffective if there are intervening quadrupole focusing elements.

**Compensation with skew quads** Alternatively, skew or rotated quadrupoles can be used to effect compensation. In general, four pairs of such coupling elements are required. In a symmetric IR (where the final focus elements are mirror symmetric about the IP), three pairs of coupling elements are sufficient.

**Transfer matrix** Define the compensation region of the lattice to include the detector solenoid and all of the compensation elements.

1. The coupling is compensated if the  $4 \times 4$  matrix ( $\mathbf{T}^{\text{through}}$ ) through the compensation region is block diagonal.
2. Vertical displacement at the IP is independent of the horizontal phase space outside the compensation region.

The second criterion imposes constraints on the matrix ( $\mathbf{T}^{\text{outside} \rightarrow \text{IP}}$ ) that propagates trajectories from outside the compensation region to the IP,  $\mathbf{T}_{31}^{\text{outside} \rightarrow \text{IP}} = \mathbf{T}_{32}^{\text{outside} \rightarrow \text{IP}} = 0$ .

Let  $\mathbf{T}_{\text{right} \rightarrow \text{IP}}$  and  $\mathbf{T}_{\text{left} \rightarrow \text{IP}}$  be the matrices that transport trajectories through the right and left halves of the compensation region to the IP. Then  $\mathbf{T}_{\text{through}} = \mathbf{T}_{\text{right} \rightarrow \text{IP}} \mathbf{J} \mathbf{T}_{\text{left} \rightarrow \text{IP}}^{-1} \mathbf{J}$ , where

$$\mathbf{J} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}$$

**Compensation with four pairs of rotated quadrupoles** It is clear that if  $\mathbf{T}_{\text{right} \rightarrow \text{IP}}$  and  $\mathbf{T}_{\text{left} \rightarrow \text{IP}}$  are block diagonal then  $\mathbf{T}_{\text{through}}$  is block diagonal and the compensation criteria are satisfied. Symplectic matrices have the property that if one off diagonal  $2 \times 2$  block is all zero, then so is the other. Therefore four independent coupling elements (skew or rotated quadrupoles) are sufficient to satisfy the compensation criteria.

**Compensation with three pairs of rotated quadrupoles** But it is not necessary that  $\mathbf{T}_{\text{right} \rightarrow \text{IP}}$  be block diagonal. Suppose that the focusing quadrupoles are deployed symmetrically about the IP and the compensating rotated quads antisymmetrically, (equal but opposite rotation angles). (Note that the radial fringe at the left end of the solenoid is of opposite sign to the fringe at the east end.) It is shown [5] that if

1.  $T_{31} = 0$
2.  $T_{32} = 0$
3.  $-T_{41}T_{12} + T_{42}T_{11} = 0$

where  $\mathbf{T} \equiv \mathbf{T}^{\text{right} \rightarrow \text{IP}}$ , then  $\mathbf{T}_{\text{through}}$  is block diagonal. The three constraints imposed on the matrix can be realized with three independent pairs of rotated quadrupoles. Independence of course implies nondegenerate betatron phase advance between the rotated quads.

In CESR (5 GeV/beam), the solenoid that is centered at the interaction point has a field of 1.5 T and length of 3.5 m. The final focus quadrupoles are rotated antisymmetrically to compensate the coupling of the solenoid. The three pairs of quadrupoles are rotated by  $\pm 4.5^\circ, \pm 6.8^\circ$  and  $\pm 14.0^\circ$ .

**Coupling matrix** The compensation conditions can be cast in terms of the C-matrix elements. According to Eqs.(36, 37), the blowup of the beam cross section is proportional to  $\bar{c}_{12}$  and the tilt of the beam is proportional to  $\bar{c}_{22}$ . The requirement of no vertical enlargement or twist at the IP implies that

$$\bar{\mathbf{C}}_{\text{IP}} = \begin{bmatrix} \bar{c}_{11} & 0 \\ \bar{c}_{21} & 0 \end{bmatrix} \quad (38)$$

And the machine is globally decoupled if at any and all points outside the insertion

$$\bar{\mathbf{C}} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \quad (39)$$

Eqs.(38, 39) are true if and only if the compensation criteria described above are satisfied.

The three pair compensation scheme can also be described in terms of the C-matrix elements. If the compensating elements are placed antisymmetrically about the IP, then by symmetry, the tilt of the beam  $\bar{c}_{22}$  at the IP is zero. Similarly, by symmetry,  $\bar{c}_{22} = \bar{c}_{11}$  at a symmetry point outside the compensation region, (such as the point diametrically opposite the IP). So, if antisymmetric pairs of rotated or skew quads are used to compensate for the solenoid, then three constraints must

be satisfied to ensure that there is no coupling of horizontal motion into the vertical plane at the IP, and that the full turn matrix evaluated everywhere outside the compensation region is block diagonal. They are

1.  $\bar{c}_{12}^* = 0$
2.  $\bar{c}_{12}^{\text{outside}} = 0$
3.  $\bar{c}_{21}^{\text{outside}} = 0$

where  $\bar{c}^{\text{outside}}$  refers to coupling parameters at the symmetry point outside of the compensation region. Then Eqs.(38, 39) are satisfied.

**Dispersion** Vertical dispersion at the IP increases the vertical beam size and is a source via the beam-beam interaction of synchrobetatron coupling. As long as there are no bending magnets and zero horizontal dispersion inside of the compensation region, then the compensation schemes described above will generate no vertical dispersion at the IP. But if there is horizontal dispersion inside of the compensation region then additional constraints will be required.

## References

- [1] D. Edwards, L. Teng, PAC 73, 885
- [2] M. Billing, Cornell Report CBN 85-2 (1985)
- [3] D. Sagan, D. Rubin, Cornell Report CBN 96-20 (1996)
- [4] S. Peggs, PA 12 (1982) 219
- [5] D. Rubin, Cornell CON 84-4 (1984)
- [6] D. Rubin, Cornell CON 84-5 (1984)
- [7] P. Bagley, D. Rubin, PAC 87, 1301

### 4.7.5 Modeling and Control of Storage Rings Using Orbit Measurements

*J. Safranek, SLAC*

Due to magnet strength and misalignment errors, a real storage ring differs from its design. To adjust the real ring to achieve the design performance, one can apply methods which are based on measured orbits. Descriptions here are for storage rings, but many of the techniques are also useful for linacs. We first describe methods using closed orbits. Then methods using turn-by-turn measurements of betatron oscillations are described. Application of a Model Independent Analysis to storage rings will be discussed in Sec.4.8.2.

**Closed orbit analysis** When the strength of a steering magnet is changed to deflect the beam by an angle  $\theta$ , the resulting closed orbit shift at

a BPM is given by Eq.(1), Sec.4.7.1. Averaging the data for many turns, these closed orbit shifts can often be measured with accuracies of  $1 \mu\text{m}$  or better, so they provide a wealth of precise optics information.

**Beam-based alignment** In this application,  $\theta$  is generated by changing the strength of a quadrupole rather than a steering magnet. If the initial closed orbit of the beam relative to the magnetic center of the quad is  $x$ , then changing the quad strength by  $\Delta KL$  gives  $\theta = x\Delta KL$ . Beam-based alignment simply entails steering the beam toward the quad center by minimizing the measured orbit shift associated with changing the quad strength [1, 2].

Beam-based alignment is useful because BPMs are more accurate for measuring shifts in the closed orbit, but quads are more accurate for measuring the absolute orbit. The typical alignment error of a quad magnetic center is  $200 \mu\text{m}$ , while the typical alignment error of a BPM electrical center is  $500 \mu\text{m}$ . Furthermore, centering the beam through in the quads avoids orbital oscillation due to power supply ripple in the quad.

**Orbit response matrix analysis** In this application, the closed orbit shifts are considered a response matrix  $\mathbf{A}$  [Eq.(1) of Sec.4.7.3]. Given the design gradients of all quads, one can calculate the expected response matrix.

Conversely, given the measured response matrix from the real accelerator, one can calculate the the real gradients in each of the quads [3, 4]. This is done by finding the computer model of the storage ring that best reproduces the measured response matrix. The model quad gradients are varied to minimize the  $\chi^2$  difference between the model and measured response matrices. The steering magnet calibrations (the  $\theta$ 's) and BPM gains are also varied in the  $\chi^2$  fit.

The accuracy with which the fit parameters can be determined depends on the particulars of a storage ring, particularly on the ratio of BPM number to quad number and on the BPM noise. The example for the NSLS storage ring is shown in Tab.1.

In some cases the orbit response matrix data is insufficient to accurately constrain the fit quadrupole gradients, and the  $\chi^2$  fit either does not converge or results in unrealistic fit quadrupole strengths. In such cases it has been shown that modified  $\chi^2$  fit, which includes constraints on the quadrupole gradient variation,

## Sec.4.7: OPERATION OF CIRCULAR ACCELERATORS

Table 1: The rms uncertainty in the fit parameters due to noise in the orbit response matrix measurement. Also included is the resulting uncertainty in the  $\beta$ -functions around the ring.

| parameter                    | rms uncertainty |
|------------------------------|-----------------|
| quadrupole gradients         | .04%            |
| BPM gains                    | .5%             |
| steering-magnet calibrations | .5%             |
| $\beta$ -functions           | .08%            |

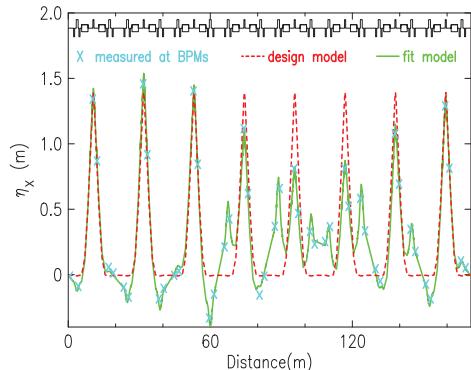


Figure 1: The measured horizontal dispersion in the NSLS X-Ray Ring compared to the design model and to the model fit to the response matrix.

yields a fit optics that reliably reproduces the actual accelerator optics [4].

Often there are significant discrepancies between the measured and design optics before applying response-matrix analysis. The model fit to the response matrix greatly improves prediction of the measured dispersion. See e.g. Fig.1, where the response matrix analysis revealed that the biggest source of optics distortion was horizontal orbit offsets in sextupoles.

Fig.2 shows the result of the response matrix analysis for ALS [5]. The 24 QD magnets have the same gradients by design, but the fit model indicated large variations, which were subsequently confirmed by power supply current measurements.

Once the source of optics distortions are understood, the gradient errors can be eliminated or compensated to restore the design optics. Fig.3 shows the NSLS VUV Ring optics fit to the orbit response matrix that had been in operation. The  $\beta$ -function distortions were associated with the

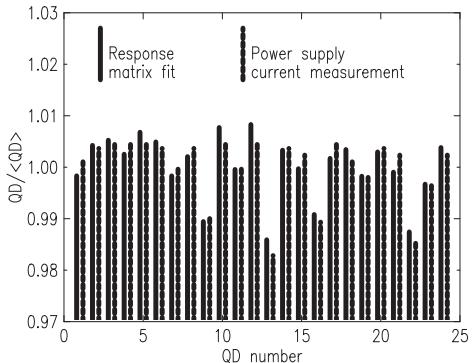


Figure 2: The gradient variations in ALS predicted by fitting the response matrix.

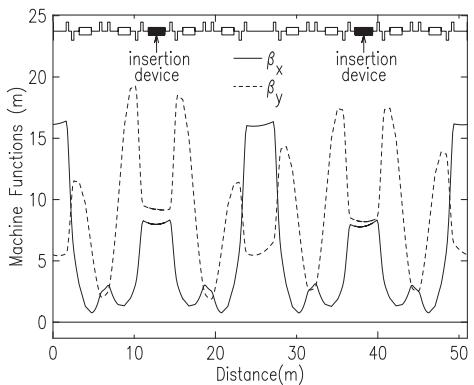


Figure 3: The VUV Ring optics before correction.

focusing in the insertion devices. With this understanding, changes in quadrupole gradients could be derived to best restore the design optics. Fig.4 shows the optics fit to the response matrix after these changes were implemented.

The orbit response matrix method can be extended to include the coupling terms such as the orbit shift in  $y$  when steering in  $x$ . These coupling terms determine the roll misalignments of the quads, steering magnets and BPMs (Sec.4.7.2).

Coupling correction (Sec.4.7.4) As an alternative to using the coupling terms in the orbit response matrix for finding the rolls of the quads, one can find instead the strengths of skew quad correction magnets that best cancel the coupling terms. The coupling of closed orbit shifts from the  $x$ - to  $y$ -plane is not quite the same as the coupling of betatron oscillations; however, it is shown

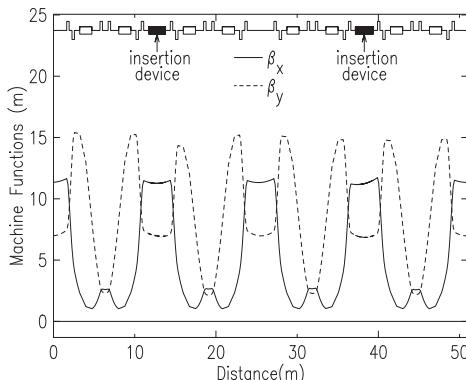


Figure 4: The VUV Ring optics with the design periodicity restored.

experimentally [6] that reducing the former helps reduce the latter.

Minimizing the vertical emittance requires minimizing both coupling and vertical dispersion. The problem can be reduced to linear algebra [6],

$$\mathbf{M}\mathbf{K}_s = -\mathbf{V} \quad (1)$$

where the vector  $\mathbf{V}$  contains all the measured shifts in vertical orbit with horizontal steering magnets (total number of elements equal to the number of horizontal steering magnets times the number of vertical BPMs) in addition to the measured vertical dispersion (an additional number of elements equal to the number of vertical BPMs). The matrix  $\mathbf{M}$  is the measured change in  $\mathbf{V}$  with each of the skew quads. Eq.(1) is solved for the desired strengths of the skew quads,  $\mathbf{K}_s$ . The relative weighting of coupling and vertical dispersion correction can be adjusted empirically to give the best reduction of vertical emittance.

Fig.5 shows the measured vertical beam size before and after implementing this coupling correction algorithm in the NSLS X-Ray Ring. The measured ratio of vertical to horizontal emittance was reduced by a factor of  $\sim 18$  to  $< 0.1\%$ .

**Turn-by-turn orbit analysis Phase and  $\beta$ -function measurements** The betatron oscillations of the beam about the closed orbit measured turn by turn at a BPM are

$$x_{b,n} = a\sqrt{\beta_b} \cos(2\pi n\nu + \phi_b) \quad (2)$$

where  $n$  is the turn number and  $a$  is the oscillation amplitude. A harmonic analysis of many consecutive turns of digitized betatron oscillation data can yield an accurate measure of  $\phi_b$ , the betatron

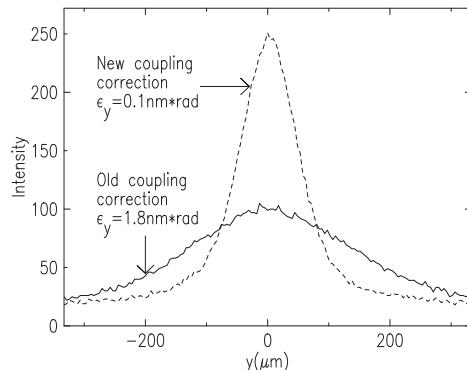


Figure 5: The measured vertical beam size in the NSLS X-Ray Ring before and after coupling correction.

phase at the BPMs [7]. An example of such measurements at LEP is shown in Sec.7.5.3, and led to a measure of the reactive transverse impedance.

The harmonic analysis alone does not provide an accurate measurement of  $\beta_b$  because the amplitude information is corrupted by BPM gain error. The  $\beta$ -function, however, is the inverse of the derivative of the betatron phase (Sec.2.1.1), so the measured betatron phases at consecutive BPMs can be used to determine the  $\beta$ -functions [9].

**Coupling correction (Sec.4.7.4)** Turn-by-turn measurements are useful in characterizing the coupling in two ways: by exciting the beam at its two normal mode frequencies, or by giving an impulse kick in  $x$  and monitoring the coupling into  $y$ .

With the normal mode technique the beam is excited resonantly at either one of the two transverse modes [10]. One mode is primarily horizontal and one primarily vertical, but both will show some excitation in each plane due to coupling. Turn-by-turn digitization of the data provides information that can be used to set the skew quads to best minimize the coupling.

With the impulse method the beam tunes are adjusted to the coupling resonance, and the beam is kicked horizontally [11]. Due to coupling, the oscillations shift back and forth between  $x$  and  $y$ . Maximizing the period of this beating minimizes the coupling.

### References

- [1] P. Röjsel, NIM A343 (1994) 364

## Sec.4.7: OPERATION OF CIRCULAR ACCELERATORS

- [2] A. Wolski, F. Zimmermann, CERN-AB-2004-010, LBNL Report-54360, ATF-03-08 (2004)
- [3] J. Safranek, NIM A388 (1997) 27
- [4] ICFA Beam Dynamics Newsletter, No. 44 (2007) p. 43, [http://icfa-usa.jlab.org/archive/newsletter/icfa\\_bd\\_nl\\_44.pdf](http://icfa-usa.jlab.org/archive/newsletter/icfa_bd_nl_44.pdf)
- [5] D. Robin et al, EPAC 96, 971
- [6] J. Safranek, S. Krinsky, AIP Proc. 315 (1994) 163
- [7] J. Borer et al, EPAC 92, 1082
- [8] D. Brandt et al, PAC 95, 570
- [9] P. Castro et al, PAC 93, 2103
- [10] P. Bagley, D. Rubin, PAC 89, 874
- [11] P.J. Bryant et al, PAC 77, 1440

### 4.7.6 Emittance Dilution Effects [1]

*M. Syphers, FNAL*

Important sources of emittance dilution in circular accelerators can be grouped into two categories. The first type, mismatch errors, occurs during the transfer of beam from one accelerator to another. Impulses, such as from kicker magnets, can also generate dilution in the same manner. The second type is from scattering and diffusion processes, such as intrabeam scattering (Sec.2.4.12), beam-gas scattering (Sec.3.3.2), power supply noise, etc. Due to synchrotron radiation damping (Sec.3.1.4.2), the beam emittance in electron synchrotrons damps to an equilibrium value. Proton synchrotrons are less forgiving, however, and the preservation of small emittance is often a serious issue. See Sec.4.3 for emittance dilution effects in linacs.

#### 4.7.6.1 Injection mismatch

If a beam is injected into a circular accelerator away from the closed orbit, or if the incoming beam ellipse is not matched to the periodic Courant-Snyder parameters of the accelerator, the beam would coherently oscillate and tumble about the closed orbit, thus occupying more phase space area than desired. Because real accelerator components have some degree of nonlinear fields present, the motion of the particles will decohere (Sec.2.3.10), generating an emittance increase. Figs.1 and 2 illustrate the effects of steering and optical mismatch errors. The phase space evolution is depicted, illustrating the effects of decoherence until an equilibrium distribution is reached, and an increase in the emittance.

We assume there is no beam damping system present. Often times, feedback systems are employed to damp injection oscillations before decoherence occurs. It is also assumed that the time-stationary beam distribution has cylindrical symmetry in the normalized phase space  $(x, \alpha x + \beta x')$  – i.e. no nearby resonances other than an amplitude dependent tune. For longitudinal phase space, this assumption implies (beam emittance)  $\ll$  (bucket area).

**Steering errors** Let  $\sigma_0$  be the incoming rms beam size at a location in the accelerator where the Courant-Snyder parameters are  $\beta_0, \alpha_0$ . If the beam enters with an initial offset  $(\Delta x, \Delta x')$ , the rms beam size after decoherence is  $\langle x^2 \rangle^{1/2}$  with

$$\langle x^2 \rangle = \sigma_0^2 + \frac{1}{2} [\Delta x^2 + (\beta_0 \Delta x' + \alpha_0 \Delta x)^2] \quad (1)$$

**Optical errors** Focusing strength errors (“ $\beta$ -mismatch”) Suppose a beamline delivers  $\beta$  and  $\alpha$  to a point in an accelerator where its periodic parameters are  $\beta_0$  and  $\alpha_0$ . The relative amplitude function mismatch,  $\Delta\beta/\beta_0$ , propagates about the accelerator (in the absence of gradient errors) according to

$$\frac{d^2}{d\phi^2} \frac{\Delta\beta}{\beta_0} + (2\nu_0)^2 \frac{\Delta\beta}{\beta_0} = -2\nu_0^2 \det \Delta\mathbf{J} \quad (2)$$

where  $\nu_0$  is the unperturbed tune,  $\phi = \psi_0/\nu_0$  is the unperturbed reduced phase advance,  $\mathbf{J} = \begin{bmatrix} \alpha & \beta \\ -\gamma & -\alpha \end{bmatrix}$ , where  $\gamma = (1 + \alpha^2)/\beta$ . The elements of  $\Delta\mathbf{J}$  are  $\Delta\beta = \beta - \beta_0$ , etc. It can be shown that  $\det \Delta\mathbf{J}$  is an invariant,

$$\det \Delta\mathbf{J} = -\frac{\left(\frac{\Delta\beta}{\beta_0}\right)^2 + \left(\Delta\alpha - \alpha_0 \frac{\Delta\beta}{\beta_0}\right)^2}{1 + \Delta\beta/\beta_0} < 0 \quad (3)$$

After decoherence the rms transverse beam size is

$$\langle x^2 \rangle = \sigma_0^2 \left( 1 + \frac{1}{2} |\det \Delta\mathbf{J}| \right) \quad (4)$$

**Dispersion function errors** If a beamline delivers a dispersion function  $D, D'$  to a point in an accelerator where its periodic dispersion function is  $D_0, D'_0$ , then off-momentum particles will see injection steering errors which after decoherence will result in a new transverse beam size

$$\langle x^2 \rangle = \sigma_0^2 + \frac{1}{2} [\Delta D^2 + (\beta_0 \Delta D' + \alpha_0 \Delta D)^2] \left( \frac{\sigma_p}{p} \right)^2 \quad (5)$$

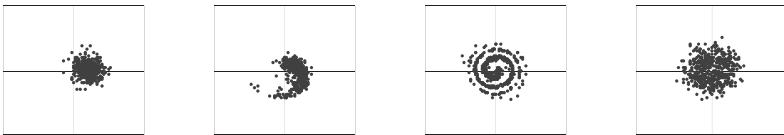


Figure 1: Steering error: Beam enters accelerator off-centered, and oscillates about the closed orbit.

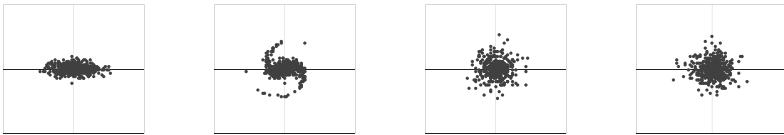


Figure 2: Amplitude function error: The same distribution as above is injected centered, but the incoming beam ellipse is mismatched with the periodic  $\beta$ -function of the accelerator (circular in this illustration).

**Longitudinal errors** For bunch areas  $\ll$  rf bucket area, similar expressions as above apply for longitudinal emittance growth due to errors in synchronized beam transfers between synchrotrons.

Energy mismatch An energy error  $\delta E$  at injection leads to a final energy spread (originally  $\sigma_E$ ),

$$\langle \Delta E^2 \rangle = \sigma_E^2 + \frac{1}{2} \delta E^2 \quad (6)$$

Phase mismatch Let  $\phi_s$  be the synchronous phase,  $V_{\text{rf}}$  the rf voltage per turn,  $\eta$  the slip factor,  $h$  the harmonic number, and  $E_0$  the synchronous energy. A phase error at injection of  $\delta\phi$  leads to a final energy spread of

$$\langle \Delta E^2 \rangle = \sigma_E^2 + \frac{\delta\phi^2}{2} \left( -\frac{E_0(v/c)eV_{\text{rf}} \cos \phi_s}{2\pi h\eta} \right) \quad (7)$$

Bucket mismatch To conserve longitudinal emittance, the buckets of the two synchrotrons should be matched. This requires  $\mathcal{K} = \mathcal{K}_0$ , where  $\mathcal{K} = (V|\cos \phi_s|)/(h|\eta|)$ .  $\mathcal{K}$  corresponds to the incoming beam and  $\mathcal{K}_0$  to the receiving synchrotron. If the two synchrotrons are mismatched then the final energy spread is

$$\langle \Delta E^2 \rangle = \left( \frac{1 + \mathcal{K}/\mathcal{K}_0}{2\sqrt{\mathcal{K}/\mathcal{K}_0}} \right) \sigma_E^2 \quad (8)$$

#### 4.7.6.2 Diffusion processes

Let  $\psi(x, x', t)$  be the particle transverse phase space density function at time  $t$ . Suppose some process is randomly altering the particles' phase space amplitude over time. Writing the density

function in terms of the individual particle emittance,  $W \equiv [x^2 + (\beta x' + \alpha x)^2]/\beta$ ,  $\psi$  satisfies the diffusion equation (see also the related Fokker-Planck equation in Sec.2.4.7)

$$\frac{\partial \psi}{\partial t} = R \frac{\partial}{\partial W} \left( W \frac{\partial \psi}{\partial W} \right) \quad (9)$$

where  $R = d\langle W \rangle / dt$ , the average being taken over the particle distribution. The rate of change of the transverse (normalized, rms) emittance would be  $d\epsilon_N / dt = \frac{1}{2}(\gamma v/c)R$ .

**Dipole noise** For a single dipole steering error randomly fluctuating each revolution of the accelerator with rms value  $\theta_{\text{rms}}$ , the emittance growth rate is

$$\frac{d\epsilon_N}{dt} = \frac{1}{2} f_0 (\gamma v/c) \beta_0 \theta_{\text{rms}}^2 \quad (10)$$

where  $\beta_0$  is the  $\beta$ -function at the location of the error, and  $f_0$  is the revolution frequency.

**Rf noise** If  $v_{\text{rms}}$  is the rms voltage noise (summed over all the cavities), uncorrelated from turn to turn, the transverse emittance growth rate is

$$\frac{d\epsilon_N}{dt} = \frac{1}{2} f_0 \gamma \langle \mathcal{H} \rangle \frac{e^2 v_{\text{rms}}^2}{(v/c)^3 E_0^2} \quad (11)$$

where  $E_0$  is the synchronous energy,  $\mathcal{H} = [D^2 + (\beta D' + \alpha D)^2]/\beta$  at the cavities.

**Coulomb scattering** If the scattering is due to small angle Coulomb interactions between the beam particles and other material in the beam chamber, then

$$\frac{d\epsilon_N}{dt} = \frac{1}{2} f_0 \langle \beta \rangle \left( \frac{13.6 \text{ MeV}}{mc^2} \right)^2 \frac{z}{\gamma(v/c)^3} \frac{\ell}{X_0} \quad (12)$$

## Sec.4.7: OPERATION OF CIRCULAR ACCELERATORS

Table 1: Relative and absolute increase in (normalized, rms) emittance due to injection errors. Here,  $\epsilon_N = (\sigma_0^2/\beta_0)(\gamma v/c)$ ,  $\epsilon_s = \sigma_E \sigma_\phi / \omega_{rf}$ .

| Error Source                                 | Relative Increase, $\Delta\epsilon/\epsilon$  | rms Emittance Increase, $\Delta\epsilon$  |
|--|---|---|
| <b>Transverse, <math>\epsilon_N</math></b>   |   |   |
| Steering Error                               | $\frac{1}{2} \frac{\Delta x^2 + (\beta_0 \Delta x' + \alpha_0 \Delta x)^2}{\sigma_0^2}$                                   | $\frac{1}{2} \frac{\Delta x^2 + (\beta_0 \Delta x' + \alpha_0 \Delta x)^2}{\beta_0} (\gamma v/c)$                                   |
| Disp. Mismatch                               | $\frac{1}{2} \frac{\Delta D^2 + (\beta_0 \Delta D' + \alpha_0 \Delta D)^2}{\sigma_0^2} \left(\frac{\sigma_p}{p}\right)^2$ | $\frac{1}{2} \frac{\Delta D^2 + (\beta_0 \Delta D' + \alpha_0 \Delta D)^2}{\beta_0} (\gamma v/c) \left(\frac{\sigma_p}{p}\right)^2$ |
| $\beta$ - Mismatch                           | $\frac{1}{2}  \det \Delta J $   | $\frac{\epsilon_N}{2}  \det \Delta J $  |
| <b>Longitudinal, <math>\epsilon_s</math></b> |   |   |
| Energy Mismatch                              | $\frac{1}{2} \left(\frac{\delta E}{\sigma_E}\right)^2$  | $\frac{1}{2} \frac{E_0}{\omega_{rf}} \sqrt{-\frac{2\pi h\eta E_0}{(v/c)eV_{rf} \cos \phi_s}} \left(\frac{\delta E}{E_0}\right)^2$   |
| Phase Mismatch                               | $\frac{1}{2} \left(\frac{\delta\phi}{\sigma_\phi}\right)^2$   | $\frac{1}{2} \frac{E_0}{\omega_{rf}} \sqrt{-\frac{(v/c)eV_{rf} \cos \phi_s}{2\pi h\eta E_0}} (\delta\phi)^2$                        |
| Bucket Mismatch                              | $\frac{1}{2} \frac{(1 - \sqrt{1 + \Delta\mathcal{K}/\mathcal{K}_0})^2}{\sqrt{1 + \Delta\mathcal{K}/\mathcal{K}_0}}$       | $\frac{\epsilon_s}{2} \frac{(1 - \sqrt{1 + \Delta\mathcal{K}/\mathcal{K}_0})^2}{\sqrt{1 + \Delta\mathcal{K}/\mathcal{K}_0}}$        |

where  $mc^2$  is the rest energy of a beam particle,  $z$  its charge, and  $X_0$  is the radiation length of the material which is distributed over a total distance  $\ell$  about the circumference. For example, a high energy proton beam scattering with the residual gas (assumed to be air) gives

$$\frac{d\epsilon_N}{dt} = \langle \beta \rangle (1.6 \times 10^{-7} / \text{s}) \frac{P[\mu\text{Torr}]}{\gamma} \quad (13)$$

**Beam-beam elastic scattering** In a collider, elastic scattering of the on-coming beam particles causes emittance growth. The emittance growth rate due to scattering at a single IP is

$$\begin{aligned} \frac{d\epsilon_N}{dt} &= \frac{1}{2} \gamma \beta^* \langle \theta_{rms}^2 \rangle \mathcal{L} \sigma_{el} \frac{1}{n_B N_B} \\ &= \frac{1}{2} \gamma \frac{f_0 N_B}{4\epsilon_N} \sigma_{el} \langle \theta_{rms}^2 \rangle \end{aligned} \quad (14)$$

where  $\beta^*$  is the  $\beta$ -function at the IP,  $\sigma_{el}$  is the elastic scattering cross section,  $\theta_{rms}$  is the rms projected elastic scattering angle. Expressions for  $\sigma_{el}$  and  $\theta_{rms}$  can be found in Secs.3.3.1,3.3.2.

**Longitudinal effects** Voltage noise in the rf system, with rms value  $v_{rms}$ , causes a growth of the longitudinal emittance at the rate

$$\frac{d\epsilon_s}{dt} = \frac{1}{2} \frac{E_0}{\omega_{rf}} f_0 \sqrt{-\frac{2\pi h\eta E_0}{(v/c)eV_{rf} \cos \phi_s}} \left(\frac{ev_{rms}}{E_0}\right)^2 \quad (15)$$

Likewise, rf phase noise causes an emittance growth rate of

$$\frac{d\epsilon_s}{dt} = \frac{1}{2} \frac{E_0}{\omega_{rf}} f_0 \sqrt{-\frac{(v/c)eV_{rf} \cos \phi_s}{2\pi h\eta E_0}} (\delta\phi_{rms})^2 \quad (16)$$

## References

- [1] D.A. Edwards, M.J. Syphers, An Introduction to the Physics of High Energy Accelerators, Wiley (1993)

### 4.7.7 Real-Time Measurement and Control of Tune, Coupling and Chromaticity

R. Jones, R. Steinbagen, CERN

Accurate control of tune, coupling and chromaticity is an integral part of modern accelerator operation, reducing beam losses associated with instabilities and preserving small beam sizes. The reliable implementation of such a system must overcome the problems related to the interdependency of these measurements. As examples; poor control of coupling will lead to tune changes which cannot be compensated by quadrupole magnets alone, while changes in chromaticity can lead to modifications in the observed tune spectra which can complicate the accurate measurement of the tune. As with most feedback systems the performance is ultimately determined by the reliability of the initial measurements. Experience has shown that in order to provide such measurements all three of these properties - tune, coupling and chromaticity - must be treated simultaneously so as to provide a robust real-time control for one or all [1, 2].

#### 4.7.7.1 Tune measurement

All betatron tune measurements are based on measuring the characteristic frequency of the transverse motion of the beam. In the simplest

case the beam is given a single kick using a powerful stripline or magnetic kicker, and allowed to oscillate freely. A position pick-up is used to measure the resulting beam motion. For stable beam conditions, the betatron tune is then usually the frequency which has the highest amplitude response in the power density spectrum obtained from this data using Fast Fourier Transform (FFT) techniques. Where powerful kickers are unavailable or when it is important to minimise the frequency range over which power is put into the beam, swept frequency “chirp excitation” or band limited white noise can be used to replace the single kick. In all these cases the measurement rate is determined by the re-triggering capability of the exciter and the damping time of the beam, typically in the Hz range and achieving resolutions of  $10^{-3} \dots 10^{-4}$  in tune.

Recent advances in detector technology and acquisition systems [3] have allowed unprecedented sensitivity to be obtained with detection of oscillations possible at the nanometre level. Such systems typically do not require any additional external excitation sources, providing excellent, continuous tune measurements from residual beam motion. This is of great advantage in hadron accelerators, where any applied excitation leads to irreversible transverse emittance blow-up.

Where the residual beam motion is not sufficient to provide clean continuous tune spectra a Phase Locked Loop (PLL) system can be implemented. In this case the exciter is fed with a sine wave from a reference oscillator. The maximum amplitude response from the beam, at the tune frequency, is obtained when the phase difference between the observed beam motion and the sinusoidally driven exciter is  $90^\circ$ . By continually adjusting the reference oscillator to meet this condition, the system automatically tracks any tune changes, resulting in a continuous tune measurement. Tracking precisions down to  $10^{-6}$  can be achieved.

For particle colliders, where beam-beam forces due to collisions produce complex coherent tune spectra, Schottky monitors measuring the incoherent tune have been successfully used to provide reliable measurements.

#### 4.7.7.2 Chromaticity measurement

The most common method of measuring the chromaticity of a circular machine is to measure the

change in betatron tune as a function of the beam energy. This is usually done by varying the rf frequency, keeping the magnetic field static. The equations are

$$\begin{aligned}\Delta Q &= (\xi Q) \frac{\Delta p}{p} = Q' \frac{\Delta p}{p} = Q' \gamma_t^2 \frac{\Delta R}{R} \\ &= Q' \left( \frac{-\gamma_t^2 \gamma^2}{\gamma^2 - \gamma_t^2} \right) \frac{\Delta f}{f},\end{aligned}$$

where  $\Delta Q$  is the change in tune,  $\Delta p/p$  the relative change in momentum,  $\Delta R/R$  the relative change in radius,  $\Delta f/f$  the relative change in rf frequency and  $Q'$  ( $\xi$ ) the (normalised) chromaticity. This method lends itself well to continuous chromaticity measurement, simply requiring the addition of a small rf modulation while tracking the tune. To avoid longitudinal emittance blow-up the modulation is typically applied either well below or well above the synchrotron tune. Fast modulation is less frequently used due to uncertainties on the bunch rigidity and its high power requirements especially for superconducting high-Q cavities.

Other methods for measuring chromaticity include its determination from the width of the tune peak, the height of the synchrotron sidebands around the main tune peak, and the phase of the head-tail motion within a bunch. Although each has been used to measure and correct for chromaticity in various machines, none has proven reliable enough for inclusion in an automatic feedback system.

#### 4.7.7.3 Coupling measurement

The amount of coupling in a circular machine is described by the complex coupling coefficient  $C^- = |C^-| e^{i\chi}$  (Sec.3.1.4.4), where  $|C^-|$  is the coupling magnitude and  $\chi$  is the coupling phase. It can be shown [4] that

$$\begin{cases} Q_1 = Q_{x,0} - \frac{1}{2}\Delta + \frac{1}{2}\sqrt{\Delta^2 + |C^-|^2} \\ Q_2 = Q_{y,0} + \frac{1}{2}\Delta - \frac{1}{2}\sqrt{\Delta^2 + |C^-|^2} \end{cases} \quad (1)$$

$$\begin{cases} \Delta\phi_1 = \phi_{1,y} - \phi_{1,x} = \chi \\ \Delta\phi_2 = \phi_{2,x} - \phi_{2,y} = \pm\pi - \chi \end{cases} \quad (2)$$

where  $Q_1$  and  $Q_2$  are the measured horizontal and vertical tunes (eigenmodes), while  $Q_{x,0}$  and  $Q_{y,0}$  are the so called “unperturbed tunes”, i.e. the tunes that would be measured in the absence of coupling.  $\Delta\phi_1$  is the phase of the vertical motion with respect to the horizontal motion and  $\Delta\phi_2$

that of horizontal with respect to the vertical motion.  $\Delta$  is the difference between the fractional part of the unperturbed tunes ( $\Delta = Q_{x,0} - Q_{y,0} - p$ , with  $p$  an integer). This can be re-written as

$$|C^-| = \frac{2\sqrt{r_1 r_2} |Q_1 - Q_2|}{(1 + r_1 r_2)},$$

$$\Delta = \frac{|Q_1 - Q_2| (1 - r_1 r_2)}{(1 + r_1 r_2)}$$

where

$$\begin{cases} r_1 = \frac{A_{1,y}}{A_{1,x}} = \sqrt{\frac{\beta_y}{\beta_x}} \cdot \frac{|C^-|}{2\nu + \Delta}, \\ r_2 = \frac{A_{2,x}}{A_{2,y}} = \sqrt{\frac{\beta_x}{\beta_y}} \cdot \frac{|C^-|}{2\nu + \Delta}, \\ \nu = \frac{1}{2} \sqrt{\Delta^2 + |C^-|^2} \end{cases}$$

with  $A_{1,x}$  and  $A_{1,y}$  the amplitudes of eigenmode Q1 in the horizontal and vertical plane respectively;  $A_{2,x}$  and  $A_{2,y}$  representing the amplitudes of eigenmode Q2 in the horizontal and vertical plane respectively;  $\beta_x$  and  $\beta_y$  the optical betas at the point of observation.

All of the coherent tune measurement techniques outlined above are detecting the eigenmodes and their amplitudes, from which the unperturbed tunes and coupling magnitude can be inferred. Using PLL or chirp excitation can, in addition, give access to the coupling phase.

#### 4.7.7.4 Real-time control of tune, coupling & chromaticity

In a first-order control approach the individual parameter deviations measured by one of the above techniques are converted into corrections that, when applied to the magnet circuits, minimise the observed errors. For most accelerators the relationship between magnetic field, required circuit current and induced tune, chromaticity and coupling changes is sufficiently linear to be cast into matrices. The correction to be applied is then typically calculated using the pseudo-inverse of these matrices (e.g. using singular value decomposition). Iterating continuously compensates for dynamic parameter deviations and residual errors that may arise due to transfer function uncertainties. However, due to intrinsic limitations such as the bandwidth of the magnetic circuits and non-linear effects such as delays and rate-limits, the stabilisation is typically not instantaneous. To cope with these effects requires a more complex form of controller to optimise temporal convergence (see e.g. [5, 6]).

This individual parameter control complexity is dwarfed by the challenge of operating parallel feedback loops on tune, chromaticity, coupling, orbit, radial position and transverse bunch-by-bunch motion. Even in a fully optimised scheme, some cross-talk is inevitable: the momentum modulation required to measure chromaticity induces tune and radial offsets that are seen by the tune, orbit and radial position feedbacks; transverse feedback, by design, minimises the very same beam oscillations required to measure the tunes. If not addressed at an early design stage, a naïve one-by-one implementation of these feedback loops can lead to serious interferences, coupling and instabilities.

There are various classic de-coupling strategies such as: diagonalisation, e.g. decoupling of horizontal and vertical planes; suppression of known cross-terms, i.e. allowing certain variations which are required for measurements; dead-bands to limit the operational ranges of one feedback in favour of another; time-scheduling between feedback actions, such as alternating tune measurements with transverse feedback operation; choosing different bandwidths for each loop.

An improved de-coupling strategy is to derive the dependent variable from the compensated feedback actuator control signal. In this case the tune feedback is operated at the maximum desired bandwidth, fully compensating radial modulation induced tune changes. Chromaticity is in turn derived and corrected from the amplitude of the actuator signal required to stabilise the tunes. Due to the finite bandwidth and gain of the feedback, the actuator signal does not typically contain the full modulation. An accurate chromaticity estimate needs to account for this and should be complemented by the demodulation of the residual tune frequency oscillation remaining on the beam. The required dispersion orbit variation and corresponding momentum mismatch need to be addressed differently. This is done by subtracting them dynamically from the orbit and radial-loop feedback reference targets. In machines running with transverse feedback systems, the tune can similarly be derived from its actuator signal while keeping beam oscillations and potential instabilities under control.

Because of the various inter-loop dependencies, it is beneficial to implement the tune, chromaticity, coupling, orbit and radial-loop

feedbacks in one global controller to minimise data exchange and synchronisation requirements.

## References

- [1] M. Minty et al, PAC 11, 1394
- [2] R. Steinhagen, CAS on Beam Diagnostics, Dourdan, France (2008) 317
- [3] M. Gasior, R. Jones, DIPAC 2005, 212
- [4] R. Jones, P. Cameron, Y. Luo, BNL-C-A/AP/204
- [5] D. C. Youla et al, IEEE Trans. on Automatic Control, Vol.21-1 (1976) 3-13 & 319-338
- [6] G. Goodwin, S. Graebe, M. Salgado, Control System Design, Prentice Hall (2000)

### 4.7.8 Measurement of Dispersion by Resonant Excitation

*D.L.Rubin, Cornell U.*

#### 4.7.8.1 Introduction

Dispersion is the dependence of the closed orbit on the beam energy. In a storage ring with horizontal but no vertical bends, the horizontal dispersion characterizes the optics. Absent misalignments and transverse coupling the first order vertical dispersion is zero. Residual vertical dispersion results from vertical kicks due to offset quadrupoles and tilted dipoles, and from coupling of horizontal dispersion via tilted quadrupoles and offset sextupoles. Vertical dispersion is a principle source of vertical emittance. Measurement of vertical dispersion is essential to identifying and correcting its sources.

The traditional technique for determining the dispersion is to measure the difference in closed orbits of beams with different energies. The energy is changed by adjusting the frequency of the rf cavities. Alternatively, we recognize that dispersion represents the coupling of longitudinal and transverse motion. This allows us to exploit the techniques developed for measuring horizontal-vertical coupling, in particular resonant excitation of the normal mode frequencies and then measurement of the relative phase and amplitude of the vertical and horizontal response at each beam position monitor [1].

#### 4.7.8.2 Formalism

We begin with an illustration of horizontal dispersion as coupling of longitudinal and horizontal motion. The linear motion is characterized by a

4X4 full turn map  $T$ . We suppose that there is a single rf cavity with matrix

$$C_{rf} = \begin{pmatrix} I & 0 \\ 0 & A \end{pmatrix}, \text{ where } A = \begin{pmatrix} 1 & 0 \\ -\frac{\omega}{c} \frac{V}{E} & 1 \end{pmatrix}$$

Then

$$T = RC_{rf} \text{ where } R = \begin{pmatrix} X & X_z \\ Z_x & Z \end{pmatrix} \quad (1)$$

Write

$$X = \begin{pmatrix} \cos \theta_x - \alpha \sin \theta_x & \beta_x \sin \theta_x \\ -\gamma_x \sin \theta_x & \cos \theta_x + \alpha \sin \theta_x \end{pmatrix} \quad (2)$$

and

$$Z = \begin{pmatrix} 1 & L\alpha_p \\ 0 & 1 \end{pmatrix}$$

Since  $\delta(\vec{\eta})$  is the closed orbit for energy offset  $\delta$  we have that

$$\begin{aligned} R \begin{pmatrix} \eta \\ \eta' \\ l \\ 1 \end{pmatrix} &= \begin{pmatrix} \eta \\ \eta' \\ l' \\ 1 \end{pmatrix} \\ &\rightarrow X \begin{pmatrix} \eta \\ \eta' \end{pmatrix} + X_z \begin{pmatrix} l \\ 1 \end{pmatrix} = \begin{pmatrix} \eta \\ \eta' \end{pmatrix} \end{aligned}$$

We know that the transverse position is independent of  $l$  as long as there is no RF. Therefore we can write that

$$\begin{aligned} X_z &= (I - X) \begin{pmatrix} 0 & \eta \\ 0 & \eta' \end{pmatrix} \\ &= \begin{pmatrix} 0 & \eta(1 - \cos \theta_x + \alpha \sin \theta_x) - \eta' \beta_x \sin \theta_x \\ 0 & \gamma \eta \sin \theta_x + \eta'(1 - \cos \theta_x - \alpha \sin \theta_x) \end{pmatrix} \quad (3) \end{aligned}$$

and using the symplecticity of  $R$  we find that

$$\begin{aligned} Z_x &= \begin{pmatrix} \eta'(1 - \cos \theta_x + \alpha \sin \theta_x) - \eta' \beta_x \sin \theta_x \\ 0 \end{pmatrix} \\ &\quad \eta(\cos \theta_x + \alpha \sin \theta_x - 1) - \eta' \beta_x \sin \theta_x \\ &\quad 0 \end{aligned}$$

Then if

$$T = \begin{pmatrix} M & m \\ n & N \end{pmatrix}$$

we find that

$$H = m + n^\dagger = \begin{pmatrix} -m_{12} \frac{\omega V}{c E} & 2\eta(1 - \cos \theta_x) \\ -m_{22} \frac{\omega V}{c E} & 2\eta'(1 - \cos \theta_x) \end{pmatrix} \quad (4)$$

The coupling matrix defined as

$$C = \frac{-H \operatorname{sgn}(\operatorname{tr}[M - N])}{\gamma \sqrt{(\operatorname{tr}[M - N])^2 + 4|H|}} \quad (5)$$

## Sec.4.7: OPERATION OF CIRCULAR ACCELERATORS

where  $\gamma^2 = 1 - |C|$ , so that

$$T = VUV^{-1} \quad (6)$$

with

$$U = \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix} \text{ and } V = \begin{pmatrix} \gamma I & C \\ -C^\dagger & \gamma I \end{pmatrix}$$

( $A$  and  $B$  are  $2 \times 2$  matrices.  $I$  is the  $2 \times 2$  identity). The determinant of  $H$  as given in Eq.(4) is

$$\begin{aligned} |H| &= -\frac{2\omega V}{cE}(1 - \cos \theta_x)(m_{12}\eta' - m_{22}\eta) \\ &= -2\frac{(2\pi Q_z)^2}{L\alpha_p}(1 - \cos \theta_x)(m_{12}\eta' - m_{22}\eta). \end{aligned}$$

$L$  is the circumference, and  $\alpha_p$  the momentum compaction. In the weak coupling limit where  $4|H| \ll (\text{tr}[M - N])^2$ , that is far from the coupling resonance,  $\gamma \sim 1$  and

$$\begin{aligned} C &\sim \frac{H}{\text{tr}[M - N]} \\ &\sim \frac{H}{2(\cos \theta_x - \cos \theta_z)} \\ &\sim \frac{1}{2(\cos \theta_x - \cos \theta_z)} \\ &\sim \begin{pmatrix} -m_{12}\frac{\omega V}{cE} & 2\eta(1 - \cos \theta_x) \\ -m_{22}\frac{\omega V}{cE} & 2\eta'(1 - \cos \theta_x) \end{pmatrix} \\ &\sim \begin{pmatrix} 0 & \eta \\ 0 & \eta' \end{pmatrix} \end{aligned} \quad (7)$$

where we have assumed that  $\cos \theta_z \sim 1$ , that is synchrotron tune near zero. We see that the dispersion can be identified with elements of the coupling matrix. A measurement of the longitudinal-horizontal and longitudinal-vertical coupling yields the dispersion.

### 4.7.8.3 Measurement of the coupling matrix

We need to relate the coupling matrix elements to the quantity that we can measure, namely the vertical and horizontal amplitude and phase of the signal modulated at the synchrotron tune, at each bpm. It is convenient to use normalized phase space coordinates. We remember that the phase space 4-vector  $\vec{x}$  is related to the normalized, normal mode representation  $\vec{w}$  according to

$$\vec{x} = VG^{-1}\bar{U}\vec{w} \quad (8)$$

where

$$U = G^{-1}\bar{U}G, \quad G = \begin{pmatrix} G_a & 0 \\ 0 & G_b \end{pmatrix} \quad (9)$$

and

$$\bar{U} = \begin{pmatrix} R(\theta_a) & 0 \\ 0 & R(\theta_b) \end{pmatrix}, \quad (10)$$

$$R(\theta) = \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix}, \quad (11)$$

and

$$G_i = \begin{pmatrix} \frac{1}{\sqrt{\beta_i}} & 0 \\ -\frac{\alpha_i}{\sqrt{\beta_i}} & \sqrt{\beta_i} \end{pmatrix} \quad (12)$$

We imagine that the a-mode is horizontal(vertical) motion, and the b-mode is synchrotron motion. The beam is resonantly excited at the synchrotron tune  $Q_s$ . Then

$$\begin{pmatrix} w_a \\ w'_a \\ w_b \\ w'_b \end{pmatrix} = w_0 \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} \quad (13)$$

where we have chosen without loss of generality to specify  $\vec{w}$  at a time when  $w'_b = 0$ . Then at any later time and at location  $i$ ,

$$\begin{pmatrix} w_a \\ w'_a \\ w_b \\ w'_b \end{pmatrix} = w_0 \begin{pmatrix} 0 \\ 0 \\ \cos(\theta - \phi_{iz}) \\ \sin(\theta - \phi_{iz}) \end{pmatrix} \quad (14)$$

where  $\theta = \omega_s t$ , and  $\omega_s$  is the synchrotron tune. From a measurement of the time dependence of the position signal at the  $i^{th}$  beam position monitor we extract the transverse amplitude and phase where

$$x_i = A_{ix} \cos(\theta - \phi_{ix}) \quad (15)$$

(or  $x \rightarrow y$ ). We can similarly write the longitudinal displacement in terms of the longitudinal amplitude and phase,

$$z_i = A_{iz} \cos(\theta - \phi_{iz}) \quad (16)$$

where  $\phi_{i(x/y)}$  is the horizontal (vertical) normal mode betatron phase advance and  $\phi_{iz}$  is the longitudinal phase advance at  $\theta = 2\pi n Q_s$ . From Equation 8 we get that

$$G\vec{x} = GVG^{-1}\vec{w}$$

$$G\vec{x} = \begin{pmatrix} \gamma & \bar{C} \\ -\bar{C}^\dagger & \gamma \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ \cos(\theta - \phi_{iz}) \\ \sin(\theta - \phi_{iz}) \end{pmatrix}$$

where  $\bar{C} = G_a C G_b^{-1}$  and then

$$\begin{aligned} & \left( \begin{array}{c} \frac{1}{\sqrt{\beta_x}} A_{ix} \cos(\theta + \phi_{ix}) \\ \sim \\ \frac{1}{\sqrt{\beta_z}} A_{iz} \cos(\theta + \phi_{iz}) \\ \sim \end{array} \right) \\ &= \left( \begin{array}{c} \bar{C}_{11} \cos(\theta - \phi_{iz}) + \bar{C}_{12} \sin(\theta - \phi_{iz}) \\ \sim \\ \gamma \cos(\theta - \phi_{iz}) \\ \gamma \sin(\theta - \phi_{iz}) \end{array} \right) \end{aligned}$$

With some rearrangement we find that

$$\begin{aligned} \bar{C}_{12} &= \sqrt{\frac{\beta_{iz}}{\beta_{ix}}} \frac{A_{ix}}{A_{iz}} \sin(\phi_{ix} - \phi_{iz}) \\ \bar{C}_{11} &= \sqrt{\frac{\beta_{iz}}{\beta_{ix}}} \frac{A_{ix}}{A_{iz}} \cos(\phi_{ix} - \phi_{iz}) \end{aligned}$$

Finally,  $C_{12} = \sqrt{\beta_a \beta_b} \bar{C}_{12}$  and then according to Equation 7

$$\begin{aligned} \eta_{x/y} &= C_{12} = \beta_{iz} \frac{A_{i(x/y)}}{A_{iz}} \sin(\phi_{i(x/y)} - \phi_{iz}) \\ &= \sqrt{\beta_{iz}} \frac{A_{i(x/y)}}{a_z} \sin(\phi_{i(x/y)} - \phi_{iz}) \end{aligned}$$

where  $A_{iz} = a_z \sqrt{\beta_{iz}}$ . Since there is no good way to measure the longitudinal phase and amplitude of the beam at each of the BPMs, we depend on the model calculation. Since the longitudinal focusing is typically very weak, the longitudinal phase advances very slowly and any errors that might arise due to discrepancy between model and measurement tend to be very small.

The longitudinal amplitude is not measured. In practice we determine the longitudinal amplitude (the amplitude of the energy oscillation) by fitting the measured  $C_{12}$  (horizontal dispersion data) to the model dispersion. The fitted amplitude is used to determine vertical dispersion.

In summary, to measure the dispersion by resonant excitation

1. Drive the beam at the synchrotron tune
2. Measure the amplitude and phase of the horizontal and vertical components of the motion at each beam position monitor.
3. Compute the longitudinal beta and phase at each BPM from the machine model.

4. Use the measured horizontal and computed longitudinal phase and amplitude to determine a quantity proportional to the horizontal dispersion and fit to the model dispersion to determine the amplitude of energy oscillations
5. From the measured vertical phase and amplitude, the computed longitudinal phase, and the fitted amplitude of energy oscillations, determine the vertical dispersion.

In the CesrTA at the Cornell Electron Storage Ring, the technique yields a measurement of the vertical dispersion with few millimeter resolution in a few seconds.

## References

- [1] D. Sagan et al, PRST-AB 3, 092801 (2000)

## 4.8 TEMPORAL AND SPATIAL CORRELATIONS IN BPM MEASUREMENTS

One may easily and rapidly take buffered BPM data pulse by pulse in linear or turn by turn in circular accelerators. These BPM readings will be correlated both temporally and spatially. Among many advantages, the derived information follows a statistical square root N rule in enhancing the measurement resolution, where N can be the number of pulses or turns and/or the number of BPMs. Depending on the accelerator and the physical quantities to be measured, taking these BPM readings can often be non-invasive.

### 4.8.1 Non-Invasive Measurement for Linacs

*J. Irwin, Y. Yan, SLAC*

The pulse-by-pulse incoming jitter can offer extractable correlated signals through use of a singular-value decomposition (SVD). Using  $M$  BPMs together has the immediate advantage of improving the resolution on the motion by  $\sqrt{M}$  [1]. The number of independent eigenvectors with eigenvalues above the noise floor determines the number of physical variables that are changing and affecting beam centroid motion. The linear space of the eigenvectors provides an orthogonal coordinate system on the linear space spanned by the underlying changes, and the time

patterns of the occurrence of these eigenvectors provide an orthogonal coordinate system on the linear space of the time patterns of the physical changes. See Sec.4.8.2 for application to storage rings.

Let  $b_p^m$  represent the reading of the  $m$ th BPM ( $m = 1$  to  $M$ ) on the  $p$ th pulse ( $p = 1$  to  $P$ ). The vector  $\vec{b}(t_p) = (b_p^1, b_p^2, \dots, b_p^M) \equiv \vec{b}_p$  will represent the complete set of readings on the  $p$ th pulse, and  $\mathbf{B} = [b_p^m]$  will be the  $P \times M$  matrix of BPM readings with  $\vec{b}_p$  as the  $p$ th row. The expectation is that

$$\begin{aligned}\vec{b}_p &= \vec{b}_0 + \sum_{s=1}^S \Delta q_p^s \partial_s \vec{b} + \sum_{r,s=1}^S \Delta q_p^r \Delta q_p^s \partial_r \partial_s \vec{b} \\ &\quad + \dots + \vec{n}_p\end{aligned}$$

Here  $\vec{b}_0$  is a constant arising from centroid or BPM offsets,  $\Delta q_p^s$  are the values of the  $s$ th variable on the  $p$ th pulse,  $\partial_s \vec{b}$  are the first order derivatives of the beam centroid due to the change in the  $s$ th variable, and  $\partial_r \partial_s \vec{b}$  are the second order changes.  $\vec{n}_p$  is random noise in each BPM on the  $p$ th pulse. Anything that is not random will occur in the sum of variables. It is useful to divide each variable change  $\Delta q_p^s$  by its rms value to get “normalized” change patterns  $\widehat{\Delta q_p^s}$ , and multiply the derivatives  $\partial_s \vec{b}$  by the same rms change to get rms shape functions,  $\vec{f}_s \equiv \Delta q_{\text{rms}}^s \partial_s \vec{b}$ . The same is done for the product variables. The  $\vec{f}_s$  all have the dimension of length and the normalized change patterns are dimensionless. The normalized time patterns can be assembled as a  $[P \times S(S+1)/2]$  variable matrix  $\widehat{\mathbf{Q}} = [\widehat{\Delta q_p^s}, \dots, \widehat{\Delta q_p^r \Delta q_p^s}, \dots]$ . And the space pattern functions  $\vec{f}_s$  can be put into an  $S(S+1)/2 \times M$  matrix  $\mathbf{F}$ . With this notation we may write our expectation for  $\mathbf{B}$  as  $\mathbf{B} = \widehat{\mathbf{Q}} \mathbf{F} + \mathbf{N}$ , where  $\mathbf{N}$  is the collection of noise vectors. If any variable shapes are known to be of negligible magnitude, they can be dropped from  $\widehat{\mathbf{Q}}$  and  $\mathbf{F}$ .

If one of the time patterns  $\widehat{\Delta q_p^s}$  is known (a suspected 60 Hz signal, or an important beam parameter has been measured) and furthermore known or suspected to be uncorrelated to the other time patterns, the corresponding shape pattern can be found by multiplying the measurements matrix  $\mathbf{B}$  by  $\widehat{\mathbf{Q}}_s^t = [\widehat{\Delta q_p^s}]^t$ . The error is only  $\sigma/\sqrt{P}$  where  $\sigma$  is the rms BPM noise.  $P$  can be very

large, the order of  $10^4$ , if the control system is capable of buffering pulse-by-pulse BPM data.

If a set of time patterns are known, and it is known or suspected that this set is not correlated to variables outside the set, then the shape patterns corresponding to all these variables can be found from

$$\mathbf{F}_{S'} = \mathbf{C}_{S'}^{-1} \widehat{\mathbf{Q}}_{S'}^t \mathbf{B} + \mathcal{O}(\frac{\sigma}{\sqrt{P}})$$

where  $\mathbf{C}_{S'} = \widehat{\mathbf{Q}}_{S'}^t \widehat{\mathbf{Q}}_{S'}$  is the correlation matrix of the variables in  $S'$ . If there was an unknown variable correlated to one or more of the time patterns in  $S'$ , the shape pattern corresponding to that variable will be mixed into the above calculation such that each deduced  $\vec{f}_s$  may contain a bit of that pattern. Even in this case it is usually a good strategy to use the known time patterns in this way and later attempt to identify and remove unknown patterns. Additional time patterns can be put into the  $\widehat{\mathbf{Q}}$  matrix without negative consequence. If the patterns are not present, the corresponding shapes will be zero, modulo a vector of magnitude  $\sigma/\sqrt{P}$ . Errors which arise if there are random errors in the time patterns are second order in the magnitude of the error, i.e. 10% random error in the time pattern gives a 1% mixing of the space patterns.

Of course some variables may be totally unknown, and some variables may be known as important, but not measured. The motions from these variables may be found by analyzing the  $\mathbf{B}$  matrix directly. What is known should first be removed, then one performs an SVD of the remainder. The SVD yields a product decomposition  $\mathbf{B} = \mathbf{U} \Lambda \mathbf{V}^t$ . Here the matrix  $\mathbf{U}$  will be a  $P \times M$  matrix of normalized (and orthogonal) time patterns,  $\mathbf{V}$  will be an  $M \times M$  matrix of normalized orthogonal (eigenvector) space patterns, and  $\Lambda$  are eigenvalues of the space patterns. Multiplying the normalized space patterns by the eigenvalues (which are the rms amplitudes of these motions) the product  $\Lambda \mathbf{V}^t$  can be interpreted as the rms space patterns. So we end up with a decomposition  $\mathbf{B} = \mathbf{U}[\Lambda \mathbf{V}^t]$  resembling the expectation  $\mathbf{B} = \widehat{\mathbf{Q}} \mathbf{F} + \mathbf{N}$ . The linear space of  $\mathbf{F}$  vectors with rms amplitude greater than  $\sigma/\sqrt{M}$  will be spanned by the  $\mathbf{V}$  vectors and the linear space of corresponding  $\widehat{\mathbf{Q}}$  vectors will be spanned by the  $\mathbf{U}$  vectors. A variety of techniques must now be employed to identify the vector direction of the physical motions.

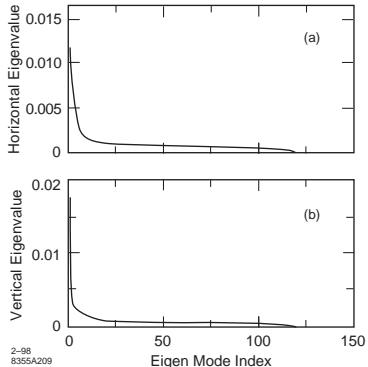


Figure 1: Typical graph of eigenvalues for several dimensions of motion plus BPM noise.

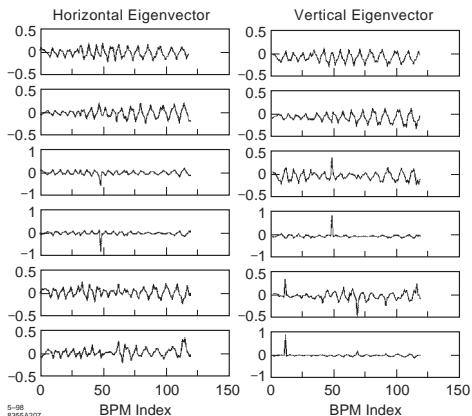


Figure 2: The leading 6 eigenvectors measured at the SLAC linac.

The first step consists of identifying the noise, and subtracting out the noise floor. A typical eigenvalue curve is shown in Fig.1. The noise will be flat and for the most part, linear. As  $P$  increases the slope of the noise floor gets smaller reaching an asymptote that corresponds to the distribution of the resolutions of the BPMs. In this way one has a direct measure of the BPM performance. The average value of the eigenvalues of the noise floor is the average BPM resolution divided by  $\sqrt{M}$ . By setting the eigenvalues of what has been identified as the noise floor to zero, and re-multiplying  $U [AV^T]$  one obtains a data set with noise removed, getting a step closer to  $\hat{Q}F$ .

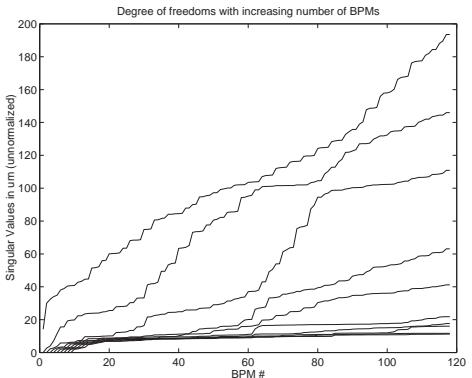


Figure 3: Degree of freedom measured at the SLAC linac (with a corrector dithered around BPM 60).

In this way we have also identified the number of degrees of freedom, i.e. the number of underlying physical variables that are changing and giving rise to changes in the beam centroid greater than or the order of about  $\sigma/\sqrt{M}$ . In this first step one should glance at the eigenvectors of the retained motion. See Fig.2. If there is a spike in these vectors, it usually indicates a bad BPM at this location. In such a case, the data from the bad BPM should be dropped from the analysis, and this step repeated.

A second step often consists of identifying initial betatron amplitudes. This can be done by using the noise-reduced data set and restricting attention to the first few BPMs. Computing an SVD of this restricted data set will typically reveal only two major eigenvectors. They can be identified as the betatron vectors and the corresponding time patterns will be the sought betatron amplitude time patterns.

The third step consists of forming an SVD for the data restricted to the first  $m'$  BPMs for every  $m'$  between 1 and  $M$ . A plot of detectable eigenvalues as a function of  $m'$ , called a “degree of freedom plot” and shown in Fig.3, reveals where in the beamline each new degree of freedom arises. By studying the SVD factorization in the region where new degrees of freedom arise, one can identify time patterns associated with this motion. There are certain limitations that arise here. The time pattern will necessarily be perpendicular to previously determined time patterns, and can differ from a physical time pattern in that regard, and the space pattern will be

## Sec.4.8: TEMPORAL AND SPATIAL CORRELATIONS IN BPM MEASUREMENTS

unique at its origin in the beamline, but may become mixed downstream with motion originating downstream.

Ideally one would make a concerted effort to measure all physical variables that could influence input conditions and have diagnostics to measure relevant beam parameters on each pulse. Then, together with the betatron amplitudes, they would form a complete set of input conditions. In addition, variables can be dithered as the data is taken, and the corresponding vectors easily isolated and identified. Combining these techniques, one can hope to limit the unknown shapes to those arising from variables in the beamline itself. By their nature these are often well localized and identifiable by the methods of the preceding paragraph.

The betatron vector, which may be determined very accurately from these procedures, can be used to study the betatron emittance pattern in the linac [which is intimately related to wake effects, BNS profiles, etc. (Sec.4.3)]. The betatron vector also can be used to differentiate shape patterns to find “kick” representations of these vectors. The kick patterns often facilitate identification and localization of sources.

## References

- [1] J. Irwin, C.X. Wang, Y.T. Yan et al, PRL 82 (1999) 1684

### 4.8.2 Invasive Measurement for $e^+, e^-$ Circular Accelerators

*J. Irwin, Y. Yan, SLAC*

An accelerator is built based on an ideal lattice design. However, its optics will not be exactly the same as the ideal lattice. On the other hand, one may be able to construct a virtual model with very much the same linear optics as the accelerator, particularly an  $e^+$  or  $e^-$  storage ring, through betatron-motion orbit measurement [1] and SVD-enhanced Least-square fitting [2].

**A complete set of data acquisition for geometric optics measurement** The linear geometric optics can be well determined if one gets 4 independent linear orbits and knows the accelerator structure. These can be obtainable from the linear mapping,  $Z^b = R^{ab}Z^a$ , written as  $R^{ab} = Z^b Z^{a-1}$ , where the  $4 \times 4$  matrix,  $Z^a = [\bar{z}_1^a, \bar{z}_2^a, \bar{z}_3^a, \bar{z}_4^a]$ , represents 4 independent linear orbits at location  $a$ , and  $R^{ab}$  is the linear map from

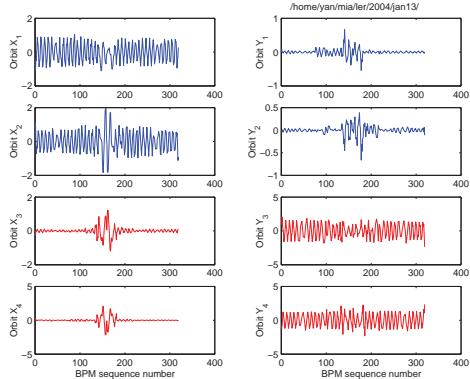


Figure 1: Four independent orbits extracted from PEP-II LER BPM buffer data. The first two orbits ( $x_1, y_1$ ) and ( $x_2, y_2$ ) are extracted from beam orbit excitation at the horizontal tune while the other two orbits ( $x_3, y_3$ ) and ( $x_4, y_4$ ) are from excitation at the vertical tune.

location  $a$  to location  $b$ . A complete set of data will allow for the extraction of the 4 independent orbits.

Unlike linacs where there is often enough incoming jitter in the beam to measure and identify betatron modes, in the  $e^\pm$  rings, to offset radiation damping, the most economic process for data acquisition is through two orthogonal resonance excitations, one at the horizontal betatron tune and the other at the vertical tune, each time buffering BPM data. Since a betatron motion has two degrees of freedom, each excitation would generate a pair of conjugate (cosine- and sine-like) betatron motion orbits from a tune-matched FFT after removing non-physical BPM data [3]. Therefore, a complete set of 4 independent linear orbits in X and Y can be extracted from the two eigen-mode excitations. A typical set of such orbits for PEP-II LER is shown in Fig.1. Note that if the storage ring is not strongly coupled, taking the sum of the squares of the cosine- and sine-like orbits from the two excitations approximately gives the horizontal and vertical betatron amplitudes in the BPM measurement space. Note that one can also perform SVD of the buffered BPM data. However, the tune-matched FFT yields more accurate extraction of the 4 independent linear orbits in the case of tune-resonance excitation.

**Fitting constraints** Since the storage ring is based on an ideal lattice, that ideal lattice can serve as the initial guess for setting its linear

magnetic components. While it is clear that all quadrupole strengths and sextupole feed-downs as well as the BPM linear gains and cross-couplings should be used as a complete set of variables, one can be flexible in choosing fitting constraints. The key to successful fitting is dealing with degeneracies which are well handled by an SVD-enhanced Least-square fitting [2].

Convenient constraints, derivable from the 4 independent orbits, include phase advances [7], Green's functions between any two BPMs [1] and for the same BPMs between any two turns [8], and eigen-plane coupling ellipse tilt-angles and axis-ratios [9]. These constraints are not necessarily independent, but for better convergence, one may include many such constraints. On the other hand, one may omit the eigen-plane coupling-ellipse tilt-angle and axis-ratio constraints in order to check that the fitted result is self-consistent [8].

**Phase advances** One can derive the orbit betatron phase at each BPM location by simply taking the arctangent of the ratio of the imaginary part to the real part of the resonance excitation FFT mode [7]. Phase advances between adjacent BPMs can then be calculated by subtraction. Note that the ratio of the imaginary part to the real part of the FFT will cancel in the linear BPM gains but not in the BPM cross-couplings. Therefore the phase advances among BPMs are repeatedly calculated during the Least Square fitting process.

**Linear Green's functions** The linear Green's function are simply the  $R_{12}^{ab}$ ,  $R_{34}^{ab}$ ,  $R_{14}^{ab}$ ,  $R_{32}^{ab}$  of the linear transfer matrix between any two BPMs labelled as  $a$  and  $b$ . They are given in the data measurement space as [1]

$$(x_1^a x_2^b - x_2^a x_1^b) / Q_{12} + (x_3^a x_4^b - x_4^a x_3^b) / Q_{34} = \mathcal{R}_{12}^{ab}$$

$$(x_1^a y_2^b - x_2^a y_1^b) / Q_{12} + (x_3^a y_4^b - x_4^a y_3^b) / Q_{34} = \mathcal{R}_{32}^{ab}$$

$$(y_1^a x_2^b - y_2^a x_1^b) / Q_{12} + (y_3^a x_4^b - y_4^a x_3^b) / Q_{34} = \mathcal{R}_{14}^{ab}$$

$$(y_1^a y_2^b - y_2^a y_1^b) / Q_{12} + (y_3^a y_4^b - y_4^a y_3^b) / Q_{34} = \mathcal{R}_{34}^{ab}$$

where  $Q_{12}$  and  $Q_{34}$  are the two invariants relating to the two resonance excitation amplitude;  $(x_1, y_1)$ ,  $(x_2, y_2)$ ,  $(x_3, y_3)$ ,  $(x_4, y_4)$  are the 4 independent linear orbits; and  $\mathcal{R}_{12}, \mathcal{R}_{32}, \mathcal{R}_{14}, \mathcal{R}_{34}$  are given as

$$\mathcal{R}_{12}^{ab} = g_x^b R_{12}^{ab} g_x^a + g_x^b R_{14}^{ab} \theta_{xy}^a + \theta_{xy}^b R_{32}^{ab} g_x^a + \theta_{xy}^b R_{34}^{ab} \theta_{xy}^a$$

$$\mathcal{R}_{32}^{ab} = g_y^b R_{32}^{ab} g_x^a + g_y^b R_{34}^{ab} \theta_{xy}^a + \theta_{yx}^b R_{12}^{ab} g_x^a + \theta_{yx}^b R_{14}^{ab} \theta_{xy}^a$$

$$\mathcal{R}_{14}^{ab} = g_y^b R_{14}^{ab} g_y^a + g_x^b R_{12}^{ab} \theta_{yx}^a + \theta_{xy}^b R_{34}^{ab} g_y^a + \theta_{xy}^b R_{32}^{ab} \theta_{yx}^a$$

$$\mathcal{R}_{34}^{ab} = g_y^b R_{34}^{ab} g_y^a + g_y^b R_{32}^{ab} \theta_{yx}^a + \theta_{yx}^b R_{14}^{ab} g_y^a + \theta_{yx}^b R_{12}^{ab} \theta_{yx}^a$$

where  $g_x$ 's,  $g_y$ 's are the BPM gains, and  $\theta_{xy}$ 's and  $\theta_{yx}$ 's are the BPM cross-coupling multipliers [1].

**Coupling ellipses** For each double-view BPM, one can trace the extracted high-resolution real-space orbits to obtain a coupling ellipse in real space for each resonance (eigen) excitation. Therefore, one can calculate coupling-ellipse tilt-angles and axis-ratios for all double-view BPMs in each of the two eigen planes [9]. For a very small  $\beta_y$  location such as the interaction point (IP) in a collider, the tilt angle of the coupling ellipse for the horizontal eigen plane is very close to the real tilt angle of the beam. One can also calculate the corresponding coupling parameters from the linear map of a lattice model. Therefore, these quantities can also be used as fitting parameters although they can be replaced with additional Green functions.

**Dispersion measurement and fitting** The linear magnet component fitting should be solely determined by geometric optics. That is, the dispersion is not part of the completeness of the constraints that determine the quads, the skews, and the sextupole feed-downs. However, if the accelerator orbit is well confined, the geometric-optics-fitted model dispersion should be close to the measurement which can be obtained from turn-by-turn BPM buffer data by exciting synchrotron oscillations. If one wishes, the same variables can be further fitted to include the dispersion constraints (with a moderate weighting) without hurting the geometric-optics fitting.

**Virtual Accelerator for Real Accelerator optics improvement** The well-fit lattice model can be viewed as a virtual accelerator matching the optics of the real accelerator [8]. As an example, shown in Fig.2 are optical characteristics plots for a well-fit PEP-II LER. Since one has the virtual accelerator in a computer, it is convenient to test for solutions that involve changing a limited number of key normal and skew quadrupoles, and/or orbit correctors (for producing sextupole bumps) to improve the virtual accelerator linear optics. One could then apply the changes to the real accelerator. These processes have been successfully applied to PEP-II LER and HER to reduce coupling effects, beta beating, and bring the working tune to near half integers for luminosity enhancement [10].

## References

- [1] J. Irwin, Y.T. Yan, EPAC 00, 151
- [2] Y.T. Yan, Y. Cai, NIM A558 (2006) 336

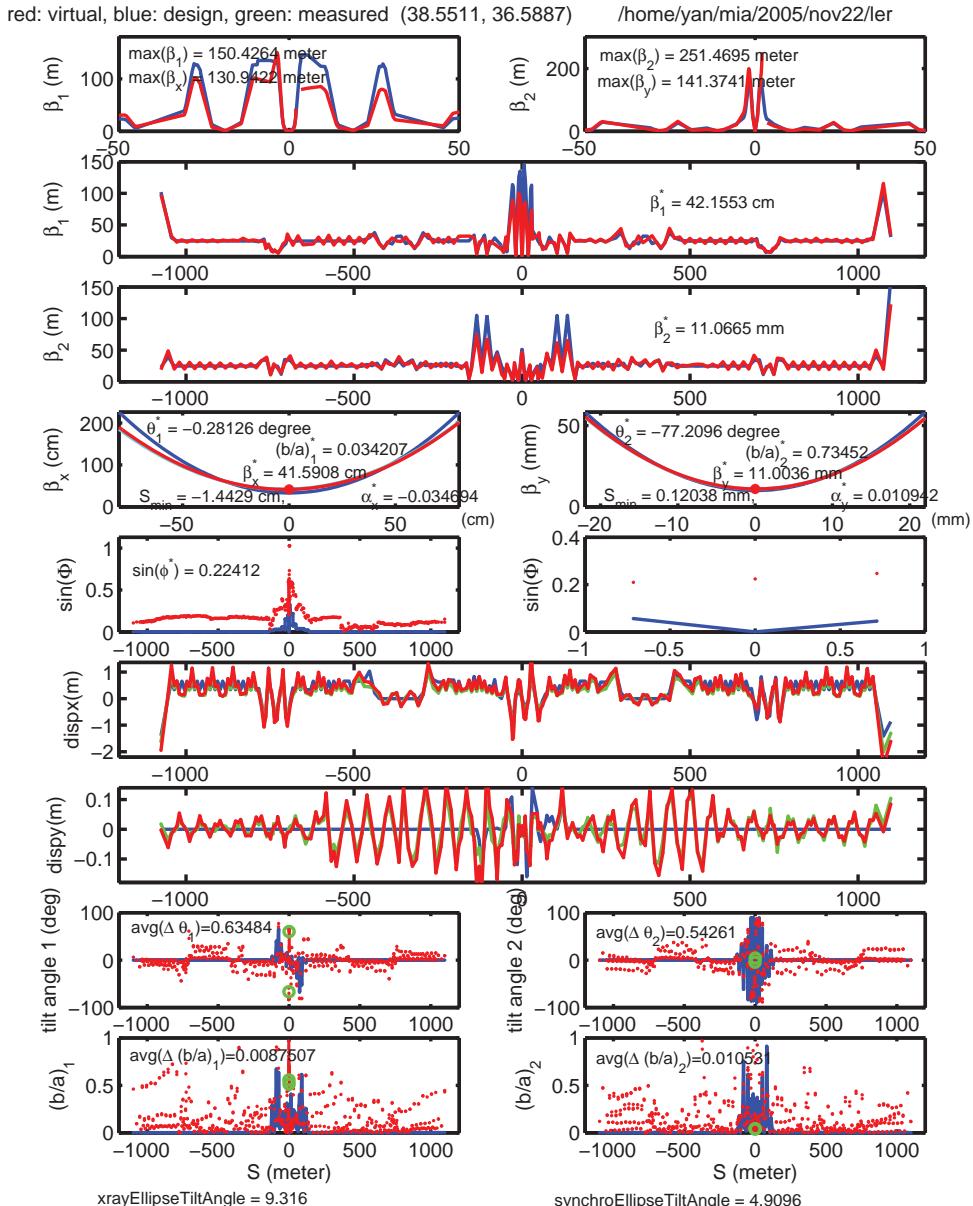


Figure 2: Plots to show a virtual accelerator linear optics characteristics (red) compared with those of the designed lattice (blue) and/or measurement (green). The top two plots show the two eigen beta functions in the vicinity of IP followed by two plots that show the  $\beta$ -functions around the ring and then the  $\beta$ -function plots at IP, which show the  $\beta^*$ 's, the waists, the coupling ellipse tilt angles ( $\theta^*$ 's), and axis ratios ( $(b/a)^*$ 's). The next two plots show the phase-space coupling angles in the generating matrix for decoupling normalization, followed by horizontal and vertical dispersion plots. The bottom 4 plots show the coupling eigen-plane ellipse tilt angles and axis ratios at all double-view BPMs.

- [3] tune-matched FFT means varying the number of turns included in the FFT so that the principle FFT mode is sharp. Alternatively, one may use a time-correlation process to identify the peak modes [4]. One may also perform an SVD to extract dominant modes for Lattice parameter measurements [5] [6].
- [4] X. Huang et al, PRST-AB 8 (2005) 064001
- [5] C.X. Wang, V. Sajaev, C.-Y. Yao PRST-AB 6, 104001 (2003)
- [6] R. Calaga et al, EPAC 04, 1541
- [7] Y.T. Yan et al, SLAC-PUB-9368 (2002)
- [8] Y.T. Yan et al, PAC 05, 3426
- [9] Y.T. Yan, Y. Cai, PAC 05, 3459
- [10] Y.T. Yan et al, SLAC-PUB-10369 (2003)

### 4.8.3 Virtual Models for Proton Circular Accelerators

*Y. Yan, SLAC*

As described in Section 4.8.2 for invasive measurement of an electron circular accelerator, one can construct a virtual model with very much the same linear optics as the real machine. Such a virtual model may be obtained for a proton circular accelerator, too. However, it requires a more elaborate procedure.

Unlike an electron circular accelerator where one can resonantly excite clean betatron motion because of synchrotron radiation damping, in a proton circular accelerator such as the LHC, one may only excite modified betatron motion by inserting an ac dipole [1] oscillating at a frequency close but not exactly equal to the horizontal or the vertical betatron tune. The modification is due to the ac dipole itself which modifies the circular accelerator as if there is an insertion of a symplectic linear map or simply a quad in one dimensional case [2]. While measuring the modified betatron motion is straightforward as one can follow a procedure similar to the one described in Sec.4.8.2, accurately subtracting the inserted linear map due to the ac dipole for establishing a virtual model can be tedious. However, taking advantage of the Green's functions from Sec. 4.8.2, one may consider using ac dipoles for beam orbit excitation but avoiding direct measurement of the modified optics quantities

To avoid measuring optics quantities modified by the ac dipole insertion, one may not directly measure the globally affected quantities such as beta functions, phase advances but the local linear Green's functions, R12, R14, R32, R34

of the transfer matrices which do not cross any of the oscillating ac dipoles. One may use two pairs of ac dipoles located as far as possible from each other. Turn on the first pair of ac dipoles. Let one dipole of this pair oscillate at near the horizontal tune while the other at near the vertical tune. Then take turn by turn beam orbit data with BPMs. Performing two focused FFTs [3] of the orbit data one for the mode near the horizontal tune and the other for the mode near the vertical tune, one would obtain two pairs of conjugate (cosine-like and sine-like) orbits for calculating local linear Green's functions between BPMs (Sec.4.8.2). Discarding any Green's function that crosses either one ac dipoles or both of them, one would get the first set of local Green's functions, none of which are modified by the inserted ac dipoles. However, this reduced set of Green's functions is not complete because there is no Green's function in the vicinity of the two active ac dipoles. To make up for the discarded local Green's functions in the area near the first pair of ac dipoles, one turns off the first pair and turns on the second pair of ac dipoles. In the same manner one would be able to get another reduced set of local Green's functions, none of which are modified by the inserted ac dipoles. Combining the two reduced sets of local Green's functions, one has a complete set of Green's functions that covers the entire circular accelerator in its original configuration without active ac dipole. This combined set of Green's functions provides a complete set of fitting constraints for deriving a virtual model by following the fitting procedures of Section 4.8.2

### References

- [1] M. Bai et al, PR E56 (1997) 6002; S. Peggs, C. Tang, BNL Report No. RHIC/AP/159 (1998)
- [2] R. Miyamoto et al, PRST-AB 11, 084002 (2008); For nonlinear effects, one may refer to R. Tomas, PRST-AB 5, 054001 (2002)
- [3] Reference [3] in Sec.4.8.2

### 4.9 TRANSITION CROSSING

*J. Wei, Michigan State U.*

During acceleration in a synchrotron, the longitudinal particle motion is non-adiabatic within a characteristic time  $\pm T_c$  near transition energy  $\gamma_T$

[1, 2],

$$T_c = \left( \frac{\pi E_s \beta_s^2 \gamma_T^3}{qeV |\cos \phi_s| \hat{\gamma} h \omega_s^2} \right)^{\frac{1}{3}} \quad (1)$$

where the subscript  $s$  denotes the synchronous value. Upon crossing, the synchronous phase ( $\phi_s$ ) typically needs to be shifted by  $\pi - 2\phi_s$  in a time much shorter than  $T_c$ . Define the momentum compaction factors  $\alpha_i$  and the phase slippage factors  $\eta_i$ ,  $i = 0, 1, 2, \dots$ , according to Sec.2.3.11 [3, 4, 5].

**Linear equations of motion** With a normalized time  $d\tau = kdt$ ,  $k = qeV |\cos \phi_s| / 2\pi h$ , the longitudinal motion is described by a Hamiltonian [5]  $H(\varphi, J; \tau) = \pm J/\beta_L$ . The action-angle variables ( $\varphi, J$ ) are related to the rf phase  $\phi$  and  $W \equiv -\Delta E/h\omega_s$  by

$$\begin{aligned} \Delta\phi &= \mp\sqrt{2J/\beta_L}(\sin \varphi + \alpha_L \cos \varphi) \\ W &= -\sqrt{2J\beta_L} \cos \varphi \end{aligned}$$

where the upper (or lower) sign is for below (or above)  $\gamma_T$ ,  $\alpha_L = -\beta'_L/2$ , and ' $'$  denotes the derivative with respect to  $\tau$ . The amplitude function  $\beta_L$  is given by

$$\begin{aligned} \frac{1}{2}\beta_L \beta''_L - \frac{1}{4}\beta'^2_L + K\beta_L^2 &= 1 \\ K &= \frac{-2\pi h^3 \omega_s^2 \eta_0}{qeV \cos \phi_s E_s \beta_s^2} \end{aligned}$$

For a constant  $\dot{\gamma}$  near transition,

$$\frac{\beta_L}{kT_c} = \frac{\pi}{3}x \left[ J_{-\frac{1}{3}}^2(y) + N_{-\frac{1}{3}}^2(y) \right] \approx 1.58 - 1.15x$$

where  $y = 2x^{3/2}/3$ ,  $x = |\Delta t|/T_c$ , and  $\Delta t$  is the time delay [5]- [7] from  $\gamma_T$ . The synchrotron frequency is  $\Omega_s = k\beta_L^{-1}$ . The maximum excursions in  $\phi$  and  $W$  are  $\hat{\phi} = \sqrt{2\gamma_L J}$  and  $\hat{W} = \sqrt{2\beta_L J}$ , where  $1 + \alpha_L^2 = \beta_L \gamma_L$ . For a bunch of rms bunch area  $S = 2\pi \langle J \rangle$ , the rms phase and momentum deviations at  $\gamma_T$  are  $\hat{\sigma}_\phi = 0.52(S/kT_c)^{1/2}$  and  $\hat{\sigma}_\delta = 0.71h\omega_s(kT_cS)^{1/2}/E_s \beta_s^2$ .

**Single-particle effects** Single-particle effects include mismatching to the accelerating rf bucket, coupling to transverse motion, [8] and various kinds of mis-timing in a time comparable to  $T_c$ .

Emittance growth due to chromatic nonlinearities (Johnsen effect [3]) is [5]

$$\frac{\Delta S}{S} \approx \begin{cases} 0.76 \frac{T_{nl}}{T_c}, & \text{for } T_{nl} \ll T_c; \\ e^{\frac{4}{3}} \left( \frac{T_{nl}}{T_c} \right)^{3/2} - 1, & \text{for } T_{nl} \geq T_c, \end{cases}$$

where the total nonlinear time  $\pm T_{nl}$  is given [5, 9, 10] by

$$T_{nl} = \left| (\alpha_1 + \frac{3\beta_s^2}{2}) \right| \frac{\sqrt{6}\hat{\sigma}_\delta \gamma_T}{\dot{\gamma}}$$

This effect was experimentally observed, [9, 11, 12] and  $\alpha_1$  was obtained by measuring the synchrotron frequency or minimum-loss timing as a function of the beam radial position. Reducing the chromatic nonlinearity using sextupole families was proposed [7] and demonstrated. [9, 12]

**Umst  tter effects** Transverse space-charge force changes the tune of each individual particle, making  $\gamma_T$  dependent on the azimuthal beam density. The amount of subsequent mismatch is inversely proportional to  $\beta_s \gamma_s^2$ , and usually negligible if  $\gamma_T$  is much higher than the injection energy.

**Multi-particle mismatch** Emittance growth due to bunch mismatch [13] under a reactive impedance  $Z_\parallel$  at the bunch frequency is proportional to the ratio of the beam-induced force to the accelerating force,

$$\frac{\Delta S}{S} \approx \frac{h\hat{I}|Z_\parallel/n|}{3V|\cos \phi_s| \hat{\sigma}_\phi^2} \quad (2)$$

where  $\hat{I}$  is the peak current at  $\gamma_T$ . Eq. 2 is valid exactly for a parabolic distribution [14] under the space charge force.

A longitudinal resistive impedance  $\mathcal{R}$  at bunch frequency causes energy dissipation, shifting the synchronous phase by  $\Delta\phi_s \approx \hat{I}\mathcal{R}/V|\cos \phi_s|$  while producing a growth

$$\frac{\Delta S}{S} \approx \frac{\hat{I}\mathcal{R}}{\sqrt{6}V|\cos \phi_s| \hat{\sigma}_\phi}$$

The change in  $\phi_s$  at transition can cause severe beam loading stress [15, 16] while the rf cavity tuning system changes the sign of the reactive beam loading compensation.

**Instabilities** A capacitive (or inductive) longitudinal coupling impedance  $Z_\parallel$  at a broad-band frequency will cause a microwave instability [14, 17, 18] during a time  $T_{mw} \approx 1.37(D_\parallel - 1)T_c$  after (or before) transition if

$$D_\parallel \approx \frac{4h\hat{I}|Z_\parallel/n|}{9V|\cos \phi_s| \hat{\sigma}_\phi^2} \geq 1 \quad (3)$$

Eq. 3 is valid exactly for a parabolic distribution under negative-mass instability above  $\gamma_T$ . A resistive longitudinal impedance may cause instability

both below and above  $\gamma_T$ . Microwave instability near  $\gamma_T$  has been experimentally observed [19] and simulated.

The transverse microwave instability [20] threshold at  $\gamma_T$  is

$$D_{\perp} \approx \frac{1.5\hat{\sigma}_{\delta}\beta_Z}{b} D_{\parallel} \geq 1$$

where  $\beta_Z$  is the average  $\beta$  function at the impedance location, and  $b$  is the beam pipe radius.

When the beam stays near  $\gamma_T$  for a relatively long time, longitudinal head-tail [21] and other slow-growing instabilities may also occur.

**Simulations** Macro-particle method [14, 17, 22] has been used to construct beam-induced forces in both the space and frequency domain. For a given numerical accuracy, the number of macro particles needed to simulate a reactive (or resistive) coupling is proportional to the cubic (or linear) power of the highest frequency considered [23].

**Transition jump** A  $\gamma_T$ -jump [13, 24] has been demonstrated on many machines to improve crossing efficiency by effectively increasing the crossing rate. Without varying the tunes, [25] a sudden change of  $\gamma_T$  is achieved by pulsing quadrupoles, often grouped in  $\pi$ -doublets, at locations of high dispersion. In order to minimize optical distortion and chromatic nonlinearity enhancement, [12] “matched, first-order” schemes [26] have been adopted for recently proposed accelerators incorporating two families of quadrupoles at regions of different dispersion. For a maximum allowable fractional growth of bunch area  $G_S \equiv \Delta S/S$ , the minimum size  $\Delta\gamma_T$  and speed  $|\dot{\gamma}_T|$  of  $\gamma_T$  jump is [5, 28]

$$\Delta\gamma_T > 2\dot{\gamma}T_{nl}, \quad \frac{|\dot{\gamma} - \dot{\gamma}_T|}{\dot{\gamma}} > \left( \frac{0.76 T_{nl}}{G_S T_c} \right)^{6/5}$$

to compensate for chromatic nonlinear effect,

$$\Delta\gamma_T > \frac{31E_s\beta_s^2qeV\gamma_T^3}{h^{1/3}|\cos\phi_s|^{1/3}\omega_s^2S^2} \left( \frac{\bar{I}|Z_{\parallel}/n|}{G_S V} \right)^{4/3}$$

$$\frac{|\dot{\gamma} - \dot{\gamma}_T|}{\dot{\gamma}} > \left( \frac{2hG_S\hat{I}|Z_{\parallel}/n|}{V|\cos\phi_s|\hat{\phi}^2} \right)^2$$

for self-field mismatch, [27, 28] and [28, 29]

$$\Delta\gamma_T > \frac{46E_s\beta_s^2qeV\gamma_T^3}{h^{1/3}|\cos\phi_s|^{1/3}\omega_s^2S^2} \left( \frac{\bar{I}|Z_{\parallel}/n|}{V} \right)^{4/3}$$

$$\frac{|\dot{\gamma} - \dot{\gamma}_T|}{\dot{\gamma}} > \left( \frac{8h\hat{I}|Z_{\parallel}/n|}{3V|\cos\phi_s|\hat{\phi}^2} \right)^2$$

for microwave instability, where  $\hat{I}$  and  $\hat{\phi}$  are values in the absence of jump,  $\bar{I} = N_0qe\omega_s/2\pi$ .

**Other compensation methods** Other methods attempted or proposed include (a) minimizing the impedance at  $\gamma_T$  by adding reactive loading [30] (also see Sec.4.15.1), (b) rf system feedback, (c) avoiding phase jump [7] by continuously varying  $V$  and  $\phi_s$ , (d) rf manipulation [7] to eliminate bunch-length oscillation, (e) artificial blow-up [7] of the longitudinal emittance, (f) reducing rf voltage [5] to alleviate chromatic effects, (g) temporarily changing the orbit circumference using programmed  $V$  and  $\phi_s$  [5], (h) using a flattened rf wave [31] to reduce  $\hat{\sigma}_{\delta}$  and  $\hat{I}$  and to provide equal acceleration for all the particles near  $\gamma_T$ .

Methods to avoid transition include (a) raising injection energy (b) reducing  $\gamma_T$  along with transverse tunes (c) creating a large or imaginary  $\gamma_T$  by either using negative bends [32] or by introducing negative dispersion in bending regions using either the harmonic approaches (e.g. enhancing higher-order Fourier component of the momentum compaction choosing the horizontal tune just below the number of lattice superperiod [33]) or the modular approaches (e.g. using a  $\pi$ -straight section to link two FODO cells containing dipoles [34]). Issues include possible poor packing factor, tunability, and dynamic aperture, large dispersion and  $\beta$ -function oscillations, and excessive number of magnet families [35]. The operating J-PARC Main Ring’s imaginary- $\gamma_T$  lattice uses supercell modules containing three FODO-cells of  $90^\circ$  phase advance, omitting dipoles in the middle FODO-cell at the high positive-dispersion location [36].

**Applications** Operating storage rings under a quasi-isochronous condition (very small  $\alpha_0$ ) has been proposed to achieve very short bunches for free electron drivers, synchrotron light sources, next generation  $e^+e^-$  colliders, and muon colliders. These designs require both an accurate control of  $\alpha_1$  to provide the necessary momentum acceptance ( $\sim \alpha_1^{-1}$ ) [37], and effective ways to damp instabilities. Obtaining short bunches by extracting near  $\gamma_T$  has also been proposed for a proton driver of muon collider and for proton-driven plasma wakefield acceleration.

## References

- [1] N.M. Blachman, E.D. Courant, RSI 20 (1949) 596
- [2] E.D. Courant, H.S. Snyder, Ann.Phys. 3 (1958) 1
- [3] K. Johnsen, CERN Symposium on High-Energy Acc., 1 (1956) 106

#### Sec.4.10: RF GYMNASTICS IN A SYNCHROTRON

- [4] K. Takayama, PA 14 (1984) 201
- [5] J. Wei, Ph.D. dissertation, Stony Brook (1990); rev. 1994
- [6] J.C. Herrera, PA 3 (1972) 49
- [7] A. Sørensen, PA 6 (1975) 141
- [8] H.G. Hereward, A. Sørensen, MPS/Int. DL 68-3 (CERN, 1968)
- [9] P. Faugeras et al, IEEE Trans.Nucl.Sci.NS-26 (1979) 3577
- [10] S.Y. Lee, J. Wei, EPAC 88, 764
- [11] J.E. Griffin et al, IEEE Trans.Nucl.Sci.NS-30 (1983) 2630; I. Kourbanis et al, PAC 91, 111; E. Ciapala et al, IEEE Trans.Nucl.Sci.NS-26 (1979) 3571
- [12] J. Wei et al, EPAC 94, 976; J. Wei et al, PAC 95, 3334
- [13] A. Sørensen, 6th Int.Conf.High-Energy Acc. (1967) 474
- [14] J. Wei, S.Y. Lee, PA 28 (1990) 77
- [15] D. Boussard, IEEE Trans.Nucl.Sci.NS-32, 1852 (1985)
- [16] S.Y. Lee, J. Wei, PAC 89, 1169
- [17] W.W. Lee, L.C. Teng, 8th Int.Conf.High-Energy Acc. (1971) 327
- [18] S.Y. Lee, J.M. Wang, IEEE Trans.Nucl.Sci.NS-52 (1985) 2323; J. Wei, S.Y. Lee, Workshop on RHIC Performance, BNL-41604 (1988) 339
- [19] K. Takayama et al, PRL 78 (1997) 871
- [20] A.W. Chao, Physics of Collective Beam Instabilities in High Energy Accelerators, Wiley (1993) p.263
- [21] H. G. Hereward, RL-74-062, EPIC/MC/48, Rutherford Lab (1974); D. Boussard, T. Lin-necar, EPAC 90, 1560
- [22] J.A. MacLachlan, FN-446 (FermiLab, 1987)
- [23] J. Wei, S.Y. Lee, A.G. Ruggiero, PA 24 (1989) 211
- [24] W. Hardt and D. Möhl, CERN ISR-300/GS/69-16 (1969); D. Möhl, CERN-ISR/300/GS/69-62 (1969); W. Hardt et al, 8th Int. Conf. High Energy Acc. (1971) 329; W.W. Lee, L.C. Teng, IEEE Trans.Nucl.Sci.NS-18 (1971) 1057; W. Hardt, 9th Int. Conf. High Energy Acc. (1974) 434; W. Merz et al, PAC87, 1343; W.K. van Asselt et al, PAC95, 3022
- [25] L.C. Teng, FN-207/400 (1970)
- [26] T. Risselada, CERN PS/90-15 (1990); A. Bogacz et al, Fermilab III Instabilities Workshop (1990) 177; S. Peggs, S. Tepikian, D. Trbojevic, PAC 93, 168
- [27] D. Möhl, PEP Note-41 (1973)
- [28] J. Wei, S. Peggs, EPAC 94, 973
- [29] S.Y. Lee, K.Y. Ng, Fermi Lab III Instabilities Workshop (1990) 170
- [30] R.J. Briggs, V.K. Neil, J. Nuclear Energy, Part C, 8 (1966) 255; The CERN Study Group on New Accelerators, AR/Int. SG/64-15 vol.1 (1964) 62; A.M. Sessler, V.G. Vaccaro, CERN 68-1 (1968)
- [31] J.E. Griffin, Fermilab TM-1734 (1991)
- [32] V.V. Vladimirska, E.K. Tarasov, Theoretical Problems of the Ring Accelerator, USSR Academy of Science (1955)
- [33] R. Gupta et al, IEEE Trans. Nucl. Sci. 32:2308 (1985); E.D. Courant et al, PAC 91, 2829
- [34] L.C. Teng, PA 4 (1972) 81; R.V. Servranckx et al, PAC89, 1355; D. Trbojevic, EPAC 90, 1536
- [35] G. Guignard, PAC 89, 915; S.Y. Lee et al, PR E48 (1993) 3040
- [36] Y. Ishi et al, APAC 98, 5D002
- [37] C. Pellegrini, D. Robin, NIM A301 (1991) 27; H. Bruck, IEEE Trans.Nucl.Sci. 20 (1973) 822; D. Robin et al, Micro Bunches Workshop, AIP Proc. 376 (1995) 181

#### 4.10 RF GYMNASTICS IN A SYNCHROTRON

*R. Garoby, CERN*

##### 4.10.1 Adiabaticity [1, 2]

An rf gymnastic is called “adiabatic” when the parameters of the synchrotron motion are changed slowly enough for the distribution of particles to be at equilibrium all along the process. The longitudinal emittance is preserved and the process is reversible. “Adiabaticity” is quantified by the dimensionless adiabaticity parameter:

$$\varepsilon = \frac{1}{\omega_S^2} \left| \frac{d\omega_S}{dt} \right| \quad (1)$$

Adiabatic means:  $\varepsilon \ll 1$  (typically  $\varepsilon < 0.1$ )

##### 4.10.2 Single Bunch Manipulations

These manipulations involve changing bunch length and energy spread.

**Controlled longitudinal blow-up [3]-[7]** This non-adiabatic technique is used to stabilize the beam by decreasing the longitudinal density of particles while providing a well-defined (typically parabolic) distribution with sharp edges.

The basic ingredients of this rf gymnastics are:

- (i) Main rf system ( $V_0$  at  $h_0$ ) keeping the beam bunched; and (ii) Modulated rf at high frequency ( $V_1$  at  $h_1 \gg h_0$ )

The principle of this operation is to excite longitudinal resonances inside the bunch by phase

Table 1: Choice of parameters in a controlled longitudinal blow-up.

| Parameter     | $\hat{V}_1/\hat{V}_0$                | $h_1/h_0$   | $\alpha$ (rad)       | $\omega_m/\omega_s$  | Duration                              |
|---------------|--------------------------------------|---|----------------------|--|---------------------------------------|
| Comment       | $\sim$ proportional to blow-up speed | <ul style="list-style-type: none"> <li><math>h_1</math> integer</li> <li>bunch length larger than 1 period at <math>h_1</math>, <math>l_b &gt; \frac{2\pi}{h_1\omega_0}</math></li> </ul> |                      | <ul style="list-style-type: none"> <li>high values tend to create tails</li> <li><math>\sim</math> even integer for uniform effect on all bunches (independence of <math>\theta</math>)</li> </ul> | blow up increases with time           |
| Typical range | 0.1 to 0.3                           | 10 to 60 for maximum blow-up speed  | $0.8\pi$ to $1.2\pi$ | 3 to 12  | $\geq 20 \cdot \frac{2\pi}{\omega_s}$ |

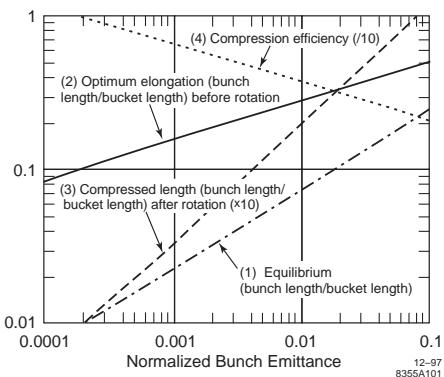


Figure 1: Equilibrium bunch length and bunch rotation parameters vs normalized bunch emittance.

modulation of  $V_1(h_1)$ , filamentation being accelerated by the nonlinearities due to  $V_1(h_1)$ . Phase modulation of  $V_1$ :

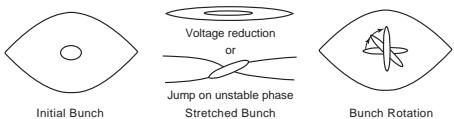
$$V_1 = \hat{V}_1 \sin(h_1\omega_0 t + \alpha \sin \omega_m t + \theta) \quad (1)$$

with  $\alpha$  is the peak phase modulation,  $\omega_m$  the modulation frequency and  $\theta$  a constant phase offset. The filamentation resonances correspond to  $\omega_m = k\omega_S$  ( $k$  integer). The effects and the typical range of parameters are given in Tab.1.

**Bunch compression [8, 9]** (i) Adiabatic variation of parameters (energy, voltage, etc ...) can reduce bunch length and increase energy spread. Curve {1} in Fig.1 gives (bunch length/bucket length) of a bunch versus its normalised emittance defined as: bunch emittance.

(ii) Non-adiabatic methods of “Bunch rotation” are faster and attain smaller length. The

principle is to stretch the bunch along the phase (time) axis, and let it rotate in the center of a high and strongly mismatched bucket.

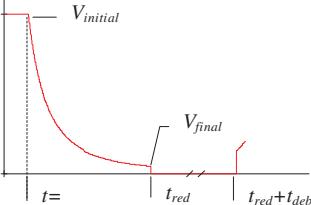
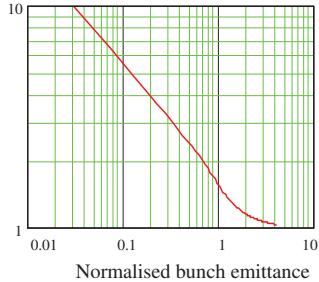


Elongation techniques include: adiabatic or non-adiabatic voltage reduction, or stretch along separatrix around the unstable phase. During bunch rotation, the bunch ends rotate slower than the centre, because of the nonlinearity of the rf waveform, giving an “S-shape” to the bunch in the phase plane and tails in the time-domain. For a given normalised emittance, the shortest bunch is approximately obtained when the bunch extremes along the time axis and the extreme along the energy axis before rotation end at the same position in time after  $\sim 1/4$  of a synchrotron period. The optimum elongation (curve {2}) and the corresponding minimum bunch length (curve {3}) are given in Fig.1, where compression efficiency (curve {4}) is defined as (adiabatic bunch length/bunch length) after optimum bunch rotation.

Application of Fig.1 to a bunch filling 1% of the bucket leads to (i) equilibrium (adiabatic) bunch length  $\sim 0.073$  of bucket length; (ii) optimum elongation before rotation  $\sim 0.27$  of bucket length; (iii) minimum length after rotation  $\sim 0.02$  of bucket length; and (iv) compression efficiency  $\sim 3.7$ .

Further improvements can be obtained by reducing nonlinearities using multiple rf harmonics.

Table 1: Choice of parameters in a debunching or rebunching operation.

| Parameter     | Voltage reduction function $V_{\text{red}}(t)$   | Duration $t_{\text{red}}$   | Final voltage $V_{\text{final}}$   |
|---------------|--|---|--|
| Comment       |  <p><math>V_{\text{initial}}</math></p> <p><math>V_{\text{final}}</math></p> <p><math>t =</math></p> <p><math>t_{\text{red}}</math></p> <p><math>t_{\text{red}} + t_{\text{deb}}</math></p> | <p>function of <math>\varepsilon</math> (adiab. param.) and <math>\omega_{s1}</math> (synch. freq. at <math>V_{\text{final}}</math>)</p> <ul style="list-style-type: none"> <li>determines emittance blow-up due to debunching</li> <li>when the bunch fills the bucket, <math>\frac{\text{debunched beam emittance}}{\text{bunched beam emittance}} = \frac{\pi}{2}</math></li> <li>the next figure shows that ratio versus normalised emittance in the bucket at <math>V_{\text{final}}</math></li> </ul> |  |
| Typical range | $\left[ 1 - \left( 1 - \sqrt{\frac{V_{\text{initial}}}{V_{\text{final}}}} \right) \frac{t}{t_{\text{red}}} \right]^2$  | $(\varepsilon < 0.1)$<br>$t_{\text{red}} \approx \frac{1}{\omega_{s1}\varepsilon}$  |  <p>Normalised bunch emittance</p> |

#### 4.10.3 Multi-Bunch Manipulations

These manipulations involve changing harmonic number and/or number of bunches.

##### Debunching-(Re)bunching [9, 10]

(i) Debunching transforms a bunched beam into a continuous beam without azimuthal modulation of density. Iso-adiabatic debunching does it with a “slow” and constant adiabaticity (Eq.(1)) reduction of the voltage from  $V_{\text{initial}}$  to  $V_{\text{final}}$ , followed by a step change from  $V_{\text{final}}$  to 0. The effects of parameters are summarized in Tab.1.

“Classical” debunching time  $t_{\text{deb}}$  is defined as the drift time required for the beginning of overlap in azimuth of particles from adjacent rectangular bunches.

$$t_{\text{deb}} = \frac{\pi - \Delta\varphi}{h\omega|\eta|\Delta p/p},$$

where  $2\Delta\phi$  and  $2\Delta p$  are the full spreads in phase and momentum.

In practice, density modulation is still too large after  $t_{\text{deb}}$  (e.g.  $t_{\text{deb}} = 0$  when the bunch fills the bucket) and more time is necessary for a good quality debunching with a small residual modulation.

(ii) (Re)bunching transforms a continuous beam into a bunched beam. Iso-adiabatic (re)bunching is the time reversed equivalent of iso-adiabatic debunching. The same rules and formulae apply.

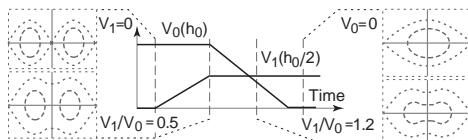
**Merging and splitting [11],[12]** Merging is used to divide the number of bunches. It can be quasi-adiabatic and preserve the type of distribution if parameters are changed slowly enough. Splitting is the reverse process.

Basic ingredients for merging pairs of bunches: two rf systems, the first one at the initial frequency ( $V_0$  at  $h_0$ ) and the second one typically at half that frequency ( $V_1$  at  $h_1 = h_0/2$ ).

Table 2: Choice of parameters in a merging or splitting operation.

| Parameter     | $V_0(h_0)$ at the beginning of merging  | Voltage functions   | Duration   |
|---------------|---|---|--|
| Comment       | <ul style="list-style-type: none"> <li>determines the normalised bunch emittance: <u>single bunch emittance bucket acceptance with <math>V_0(h_0)</math></u></li> <li>the next figure shows the ratio <math>V_1/V_0</math> at which bunches begin to merge as a function of normalised emittance</li> </ul> | linear time variations can be used, provided the normalised bunch emittance at merging is large enough ( $\sim 1/3$ ) | function of adiabaticity and tolerable blow up   |
| Typical range | <p>Normalised bunch emittance</p>   |   | $\geq 5 \cdot \frac{2\pi}{\omega_{s0}}$ where $\omega_{s0}$ is the synchrotron frequency in the bucket due to $V_0(h_0)$ at the beginning of merging |

Principle: Time variations of  $V_0$  and  $V_1$  as illustrated in the following figure translate into such an evolution in the phase plane that particles from both bunches merge into a single one. Tab.2 gives the effects of parameters in this operation.



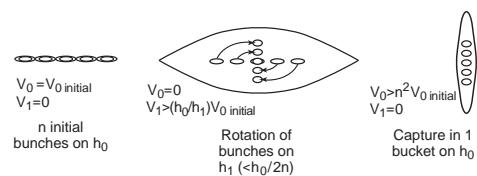
With a similar technique using three simultaneous rf harmonics, quasi-adiabatic splitting of bunches in three can be obtained without measurable emittance blow-up [12].

**Coalescing [10, 13]** Coalescing is a non-adiabatic process (large increase of the total longitudinal emittance), by which  $n$  bunches are concentrated into a single one. Rf frequency and harmonic number are the same in the initial and final states.

Basic ingredients: two rf systems, the first one at the initial frequency ( $V_0$  at  $h_0$ ) and the second one at a much smaller frequency ( $V_1$  at  $h_1 < h_0/2n$ ).

Principle: (i) initial state:  $n$  almost filled buckets with  $V_0$  initial( $h_0$ ) and  $V_1(h_1) = 0$ ;

(ii) step increase of  $V_1(h_1)$  and cancellation of  $V_0(h_0)$ , to trigger a collective rotation of the set of bunches around its centre during  $1/4$  of a turn (similar to bunch compression); (iii) suppression of  $V_1(h_1)$  and step increase to a large voltage  $V_0(h_0)$  for capturing all particles in a single bucket. See next figure.



**Slip stacking [14, 15]** Slip stacking is a non-adiabatic scheme, whereby two sets of bunches are finally concentrated into a single set.

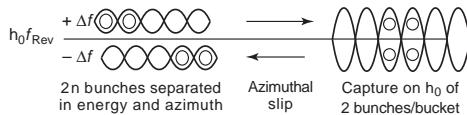
Basic ingredients: two rf systems operating with slightly different frequencies.

Principle: (i) initial state: two sets of  $n$  bunches separated in azimuth and energy, held by two rf systems (peak voltage on each system =  $V_{\text{drift}}$ ) at  $h_0 \cdot f_0 \pm \Delta f$ ; (ii) drift ("slip") of the 2 beams with respect to each other until azimuthal superposition is obtained; (iii) suppression of both rf's and step increase to a large voltage  $V_0(h_0)$  for

Table 3: Choice of parameters in a slip stacking operation.

| Parameter     | $V_{\text{drift}}$   | $2\pi\Delta f$   | $V_0$  |
|---------------|--|--|--|
| Comment       | acceptance of drifting bucket ( $A_{\text{drift}}$ ) $\gg$ bunch emittance ( $\varepsilon_l$ ) | <ul style="list-style-type: none"> <li>must be large to minimize bunch distortion and final emittance</li> <li>limited by energy acceptance</li> </ul> | large enough for the final bucket to hold both bunches             |
| Typical range | $A_{\text{drift}} \geq 2\varepsilon_l$   | $> 2\omega_s \text{ drift}$ with $\omega_s \text{ drift}/2\pi$ the synchrotron frequency in an unperturbed drifting bucket                             | $> 4V_{\text{drift}}$ for $2\pi\Delta f = 2\omega_s \text{ drift}$ |

capturing two bunches in each bucket. See figure below.

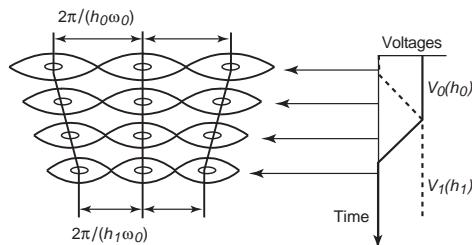


The effects of parameters on the slip stacking operation are shown in Tab.3. The emittance blow-up can be reduced with a progressive reduction of  $\Delta f$  during the final approach.

**Batch compression [16]** Batch compression is applied to squeeze a set of  $n$  bunches into a smaller fraction of the circumference. The number of bunches is unchanged. It can be quasi-adiabatic, preserving shape and emittance of bunches.

Basic ingredients:  $>2$  rf systems with a large enough tuning range.

Principle: progressively change the harmonic number of the rf experienced by the beam by “slowly” decreasing the voltage on the first harmonic  $h_0$  after having raised the voltage on  $h_1 > h_0$ . See figure below.



Limitations: reduced acceptance of the buckets displacing the edge bunches. A fraction of the circumference ( $\sim 1/5$ ) must not contain particles.

#### 4.10.4 Debunched Beam Manipulation

**Phase displacement acceleration [17, 18]** Using “Phase displacement acceleration” the energy of a debunched beam can be changed at a slow rate with a limited rf voltage. It can be quasi-adiabatic.

Basic ingredient: a single low voltage rf system with a small frequency swing (a few %).

Principle: sweeping empty rf buckets ( $V_0$  at  $h_0$ ) through a debunched beam changes the mean beam energy by:  $\Delta E = A_{\text{bucket}} f_{rf}$  where  $A_{\text{bucket}}$  is the acceptance of the sweeping bucket in eVs. Effects of parameters are summarized in Tab.1.

Limitations: small rate of acceleration and progressive degradation of the stack edges with the repeated passages of the buckets.

#### 4.10.5 Beam Manipulations with Broad-Band RF Systems

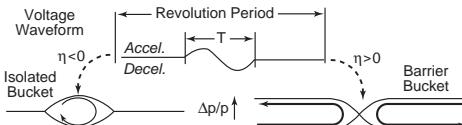
Broad-band rf systems are able to generate voltage waveforms with a large spectral content. Beyond the techniques described in the previous sections which are all applicable, they make feasible a large variety of specific manipulations.

**Barrier / Isolated bucket [10, 19, 20]** The basic building block is the beam barrier resulting from a unipolar voltage pulse.

Principle: two voltage pulses of opposite polarity with a periodicity of one turn create a single bucket. Depending on the polarity of the field and the sign of  $\eta$ , it is an “isolated” or a “barrier” bucket (see figure below). In the case of a single sine-wave (two contiguous half-sines), bucket height is given by the conventional formula (Sec.2.1.2).

Table 1: Choice of parameters in phase displacement acceleration.

| Parameter     | $V_0(h_0)$                               | $\Gamma = \sin \varphi_s$                | Duration   |
|---------------|--|--|--|
| Comment       | as large as possible to minimize blow up | as small as possible to minimize blow up | proportional to: $(A_{\text{bucket}} V_0 \Gamma)^{-1}$ |
| Typical range |  | $\Gamma \sim 0.1$                        | a few hundred sweeps                                   |



**Applications of barrier buckets [21]** Combining multiple pairs of voltage pulses and varying their amplitudes and phases as a function of time, bunches can be quasi-adiabatically accumulated or extracted from a well-defined part of a high intensity stack. Such techniques have been extensively used in the FNAL Recycler Ring [21].

Similarly, bunches can be merged or split, and their length and energy spread can be quasi-adiabatically modified [21].

Limitations: need for specific rf systems which cannot be used for fast acceleration because of their limited peak voltage resulting from the intrinsically low impedance of broad-band resonators.

## References

- [1] W.T. Weng, AIP Conf.Proc. 184, Summer Schools 1987 & 1988, p.242
- [2] M.H. Blewett (editor), CERN 77-13, 63-81
- [3] V.V. Balandin, M.B. Dyachkov, E.N. Chapochnikova, PA 35 (1991) 1
- [4] R. Cappi, R. Garoby, E. Chapochnikova, CERN/PS 92-40 (RF)
- [5] R.K. Reece et al, PAC 89, 1934
- [6] R. Cappi et al, PAC 93, 3570
- [7] J.M. Brennan et al, EPAC 94, 1897
- [8] J. Griffin et al, IEEE NS-30, No. 4 (1983) 2630
- [9] R. Garoby, CERN PS/RF/Note 93-17
- [10] J.E. Griffin, J.A. MacLachlan, Z.B. Qian, IEEE NS-30, No. 4 (1983) 2627
- [11] R. Garoby, S. Hancock, EPAC94, 282
- [12] R. Garoby, CERN/PS 98-048(RF); HEACC98 Dubna, 1998
- [13] D. Wildman et al, PAC 87, 1028
- [14] F.E. Mills, BNL Int.Rep. AADD 176 (1971)
- [15] D. Boussard, Y. Mizumachi, IEEE NS-26, No. 3 (1979) 3623
- [16] R. Garoby, IEEE NS-32, No. 5 (1985) 2332
- [17] K.N. Henrichsen, M.J. de Jonge, CERN-ISR-RF-MA/74-21
- [18] E.W. Messerschmid, CERN/ISR-TH/73-31
- [19] J.E. Griffin et al, IEEE NS-30, No. 4 (1983) 3502
- [20] M. Blaskiewicz, J.M. Brennan, EPAC 96, 2373
- [21] C.M. Bhat, RPIA 2006, FERMILAB-CONF-06-102-AD

## 4.11 ENERGY MEASUREMENT WITH ELECTRON BEAMS

*J. Seeman, SLAC*

**Single pass energy measurement** Single pass energy measurements are typically made by measuring the bend angle of a beam through a dipole magnet. However, in general, the beam is made to pass through a transport line to define a “straight” trajectory and then is made to pass into a transport section with dispersion  $D$  from several dipole magnets. If transverse beam positions relative to the nominal trajectory before and after the dipole magnet are zero, the nominal energy  $E_0$  is defined by the geometry. For a drift-dipole magnet(s)-drift geometry,

$$E_0 [\text{GeV}] = \frac{0.29979}{\theta} \int B dl \left[ \frac{\text{Tm}}{\text{rad}} \right]$$

where  $\theta$  is the measured geometrically surveyed bend angle. A 1 m long dipole with a 1 T field bends a 10 GeV beam by 30 mrad. Integrated dipole fields can be measured in the range of 0.1 to 0.5% and angles can be measured to about 0.01%. The magnet power supplies must be regulated and calibrated to provide the required resolution. If the position readings before the dipole magnet(s) or in the dispersive region are nonzero arising from either betatron oscillations or from the beam being off the standard energy, the energy must be corrected using these offsets [1, 2]. The correction  $\Delta E/E_0$  to  $E_0$  above is

$$\frac{\Delta E}{E_0} = \frac{x_3 - m_{31}x_1 - m_{32}x_2}{D_3}$$

#### Sec.4.12: RING INJECTION AND EXTRACTION

where  $x_1$  and  $x_2$  are transverse position measurements in the plane of the bend in the nondispersive transport lines ( $D_1 = D_2 = 0$ ) defining the initial trajectory and  $x_3$  a position measurement in the dispersive region with  $D_3$  nonzero.  $m_{ij}$  represent the  $R_{11}$  transport matrix elements (Sec.2.2.1) between positions  $i$  and  $j$ . It is assumed here that the positions 1 and 2 are 90 degrees apart in betatron phase, else the proper phase must be used. For example, with  $m_{ij} = 1$ ,  $D_3 = 1$  m, and  $x$  measurements on the order of 1 mm, energy corrections are typically of order 0.1%.

**Storage ring energy measurement** A stored beam passes through a series of bends and quads which create a closed orbit, approximately circular. Usually, the beam is made to pass through the centers of the quadrupoles verified by beam position measurements. The beam energy is

$$E_0 \text{ [GeV]} = \frac{0.29979}{2\pi} n \int B dl \text{ [Tm]}$$

where  $\int B dl$  is the integrated dipole field of one of the  $n$  dipole magnets around the ring. For example, a ring with 100 dipoles of length 1 m and field 0.628 T produces a beam energy of 3 GeV. Again, integrated dipole fields can be measured in the lab from 0.1 to 0.5% and magnet power supplies must match the required precision. When other dipole fields are present, corrections to the energy calculation must be made [3, 4]. Several common sources for energy corrections including typical relative contributions are: (i) Dipole correction magnets in the dispersive arcs which are used for orbit correction that have non-zero average strength (0.02%). (ii) A beam that passes off axis in the quadrupoles in the arcs, perhaps due to a changed rf frequency, from misalignments, or from tidal forces (0.05%). (iii) Temperature changes on the lengths and the pole gaps of the dipole magnets (0.02%). (iv) Extraneous currents which flow in the vacuum chambers or magnets ground paths (0.05%). (v) Energy losses from around the ring due to synchrotron radiation (0.05%). (vi) Hysteresis effects from standardization errors (0.05%).

Precise NMR measurements (0.001%) of the main dipole fields and rotating coil measurements of the relative integrated dipole fields (<0.01%) allow estimation of time varying effects on the beam energy.

Spin tune depolarization measurements using backscattered laser photons can be used to determine the beam energy to about 1 MeV abso-

lutely. See Sec.2.6.1 for the connection between the beam energy and the spin tune. See Sec.7.6 for polarimeters.

In very large storage rings the energy can vary due to these causes and others such as earth tides. Detailed reports concerning the energy calibration of the LEP storage ring, taking these effects into account, can be found in [5, 6].

Another method useful for an  $e^+e^-$  collider is to measure the ring energies when the machine is producing particles on a narrow excitation energy. For example while operating on the Y(2S) resonance, the beam energies are known to be at  $10.02326 \pm 0.00031$  or one part in about 30,000 [7, 8].

#### References

- [1] J. Seeman, AIP Proc. 249 (1992) 49
- [2] J. Frisch et al, BIW 08 (2008) 17
- [3] M. Sullivan et al, PAC 01, 3570
- [4] R. Assmann et al, Z. für Physik C, Springer-Verlag, C66 (1995) p.567
- [5] LEP Energy Working Group, CERN-EP/98-040, CERN-SL/98-012 (1998)
- [6] <http://www.cern.ch/LEPCAL/reports/reports.html>
- [7] Y. Cai et al, PRST-AB 12, 061002 (2009)
- [8] Particle Data Group, Particle Physics Booklet, July 2010, 125

#### 4.12 RING INJECTION AND EXTRACTION

*G.H. Rees, Rutherford Appleton Lab.*

##### Single turn injection with a septum and kicker

[1] Input line, lattice functions are set to match those for the ring at the septum output ( $\alpha_x = \alpha_s$ ,  $\beta_x = \beta_s$ ). Kicker unit is downstream at ( $\alpha_x = \alpha_k$ ,  $\beta_x = \beta_k$ ), and deflection  $\theta$  brings the beam on axis, after betatron phase shift,  $\mu_k$ . Input beam at output of septum, at  $(x, x', y = y' = 0)$ , requires:

$$x' = -(\alpha_s + \cot \mu_s)x/\beta_s \quad (1)$$

$$\theta = x/(\sin \mu_k \sqrt{(\beta_s \beta_k)}) \quad (2)$$

For kickers and septa see Secs.7.2.7 and 7.2.16, respectively.

**Conventional, one plane multi-turn injection with septum and orbit bump** [1] (Fig.1) Assume un-normalized, beam emittances and septum output, lattice parameters of  $(\epsilon_i, \alpha_i, \beta_i)$  for input beam and  $(\epsilon, \alpha, \beta)$  for final ring beam. The

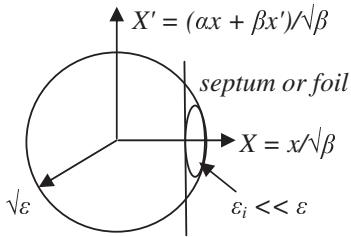


Figure 1: Optimized, 1-plane multi-turn injection. Mismatched input ellipse touches  $\sqrt{\varepsilon}$  circle once.

number of injected turns,  $n$ , is maximized by adjusting input mismatch and collapse rate of injection orbit bump. An electrostatic septum and a non-dispersive, injection straight are beneficial. Input beam at  $(x, x', y = y' = 0)$ , relative to the orbit bump, at the septum output, requires:

$$\begin{aligned}\alpha/\beta &= \alpha_i/\beta_i = -x'/x; \\ \beta/\beta_i &= \alpha/\alpha_i = (\varepsilon/\varepsilon_i)^{1/3};\end{aligned}\quad (3)$$

$$n \text{ (at low space charge)} \approx 0.5 - 0.7(\varepsilon/\varepsilon_i)$$

**Two-plane, multi-turn injection with a tilted septum and orbit bumps [2]** The number of injected turns,  $n$ , is maximized by adjusting the input mismatching, the tilt of the septum, and the collapse rates of the orbit bumps. The requirements are (using earlier symbols, but with  $y \neq 0$ ,  $y' \neq 0$  and adding subscripts  $x, y$ ):

$$\begin{aligned}\beta_x/\beta_{ix} &= \alpha_x/\alpha_{ix} = (\varepsilon_x/\varepsilon_{ix})^{1/3}; \\ \beta_y/\beta_{iy} &= \alpha_y/\alpha_{iy} = (\varepsilon_y/\varepsilon_{iy})^{1/3}; \\ \alpha_x/\beta_x &= \alpha_{ix}/\beta_{ix} = -x'/x; \\ \alpha_y/\beta_y &= \alpha_{iy}/\beta_{iy} = -y'/y \\ n &\approx (0.1 - 0.125) \frac{\varepsilon_x \varepsilon_y}{\varepsilon_{ix} \varepsilon_{iy}}\end{aligned}\quad (4)$$

**Multi-turn, charge exchange injection [1, 3]** Key items are foil, or laser or gas jet strippers, foil scattering, foil temperatures, injection mismatch and painting, and the removal of unwanted beam from the ring. After the stripper, input and circulating beams merge, and stripped electrons pass to a nearby collector. High intensity needs  $\varepsilon_{ix} \ll \varepsilon_x$ . For  $H^-$  beams, fields at the foil are set to optimize the lifetimes of excited  $H^\circ$  states.

An anti-correlated, or correlated,  $x - y$  beam distribution may be painted. Beam motion in the former is large in the  $x$ - and small in the  $y$ -plane at the start of injection, transposing to opposite

size amplitudes by the end. For the latter, beam motion is large in both planes at the start and is small by the end of injection, or vice-versa. The orbit painting bumps have to be programmed so that the beam traversals of the foil are minimized. Collapsing bumps offer advantages.

One scheme uses a fixed orbit bump, in a low dispersion region; it may have a corner foil and extra  $x, y$  orbit bumps for the painting (SNS) or a side foil with input steering for  $y'$  and an  $x$ -orbit bump (J-PARC). Another scheme (ESS, KAON) has the injection foil at the waist points of a low field, ring dipole, and uses a corner foil, a  $y$ -orbit bump and a  $p$ -momentum ramp for the  $y$ - and  $p$ -,  $x$ -painting. The dispersion at the foil, for a ramp of  $\Delta p/p$  up to  $\Delta$ , requires a  $D_x$  value:

$$D_x \leq (\sqrt{(\varepsilon_x \beta_x)} - 2\sqrt{(\varepsilon_{ix} \beta_{ix})})/\Delta \quad (5)$$

#### Charge exchange injection for FFAG rings

Most FFAGs have finite dispersion straights, too short for an injection orbit bump. An exception is a non-linear, scaling FFAG with high cell tunes,  $q_x \approx 0.78$ ,  $q_y \approx 0.36$ , which allow low dispersion straights, of a length up to  $\approx 8$  m. The non-linear,  $x - y$  ring coupling influences the distributions painted by the orbit bumps.

**Resonant injection scheme (ZGS) [4]** Involved are a fast kicker and a pulsed, bump unit with dipole, quadrupole and octupole fields. Two stable regions are formed beside an unstable fixed point. After beam is injected into the outer one, fields are set to merge the regions. Injection is repeated by re-setting conditions for the outer region, leaving injected beam in the inner region.

**Electron and Positron ( $e^\pm$ ) Injection** Fast, single-turn injection for low, electron or positron energies is similar to that for most ions. At high energies, the radiation damping plays a significant and useful injection role, as indicated in Fig.2 ( $X$  and  $X'$  are as for the Figure 1). Here,  $I$  is the mismatched injected beam,  $X_s$  the inner edge of the injection septum,  $X_{co}$  the collapsing horizontal orbit bump, and  $S$  shows the stored beam after damping. If  $X_{\sigma i}$  is the rms beam radius after injection, and  $X_\sigma$  is the rms beam radius at maximum energy, then:

$$X_s > X_{co} + 6X_{\sigma i} \text{ and } X_s \geq 10X_\sigma \quad (6)$$

#### Pulsed sextupole magnet for $e^\pm$ injection [5]

A conventional, pulsed local orbit bump, with four kickers sited along one long straight section, may be replaced by the action of a single, pulsed sextupole magnet. The parabolic waveform of the

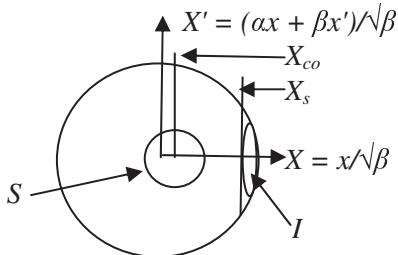


Figure 2: Multi-turn, electron-positron injection.

sextupole field allows injected beam to receive a required input deflection, while the effects on a stored beam are small due to the field reduction to zero on the sextupole axis. Key parameters are  $(x, x')$  values at the output of an injection septum, lattice functions at the septum and downstream sextupole and the horizontal betatron phase shift between the septum and the sextupole. Practical issues involve the reduction of sextupole, eddy currents, and the protection from the high power synchrotron radiation.

#### Single turn ejection with kicker and septum

This is the reverse of the single turn injection process, described earlier. A phase shift from the kicker output to septum input of  $\pi/2$  is optimum. Kickers may be electrostatic, or have a delay line or transmission line form (terminated or shorted) or be of a lumped, inductive, pure magnetic type.

**Fast, multi-turn extraction (MTE) [6]** Beam to be extracted is divided into beamlets, one on the axis and four in stable outer horizontal phase space islands. The scheme is the reverse of the ZGS, resonant injection noted earlier, but the use of sextupole and octupole fields creates four outer islands, not one. The five islands separate after sweeping through the resonance involved, which is at  $4Q_h = 25$  for the PS ring at CERN. The required separation is controlled by the  $Q_h$  value at the end of the resonance crossing and by the strength of the non-linear magnets.

The separated, circulating beamlets are moved turn by turn, beyond the septum of the extraction magnet by the use of slow and fast closed orbit bumps. In comparison with CT, the continuous transfer of beam, previously used at the PS ring, the MTE scheme has a beam loss at the septum reduced by a factor of two. The uniformity of the extracted beam is also improved, as the trapped beamlets may have similar intensities, emittances and focusing parameters. The typical,

extraction efficiencies found for MTE at the PS ring are in the range from 97 to 98%. An order of magnitude lower loss would have to be obtained if the scheme were to be used for a MW proton driver.

#### Multi-pass halo extraction by bent crystals [7]

Multi-turn halo extraction may be obtained for a high energy ring beam via a gently bent crystal at the beam edge. Halo moves out to the crystal by beam-gas scattering, beam-beam collisions or by noise. Use of a kicker may speed up the move. Positively charged, beam particles may pass into a crystal several times before channeling. Crystal dislocations and defects, and multiple scattering in a channel influence total extraction-bending efficiencies,  $E_o$ , given by:

$$E_o = E_e(1 - R_T/R_m) \exp(-s/\lambda_{db}) \quad (7)$$

where  $E_e$  is the extraction efficiency,  $R_T = p\beta c/E_c$  is the Tsyganov radius,  $E_c$  is the interatomic field where particle trajectories become unstable,  $R_m$  is the minimum bending radius of curvature,  $s$  is the crystal length and  $\lambda_{db}$  is the de-channeling length. Silicon crystals are used as they are free of large dislocations. A typical efficiency of 30% is found at Fermilab when using a 40 mm length of crystal for a 1 TeV beam.

#### Slow, third order resonant extraction [8]

Slow extraction on the third order resonance,  $3Q_x = n$ , is summarized here. Details are given in the next section. For the  $S(s) = \partial^2 B_y(s)/\partial x^2/(2B\rho)$ , sextupole excitation, the equation for  $x$  is:

$$d^2x/ds^2 + K(s)x = -S(s)(x^2 - y^2) \quad (8)$$

A normalization to  $u = x/\sqrt{\beta_x}$  and  $\varphi Q_x = \int ds/\beta_x$ , with  $r^2 = \varepsilon = u^2 + p^2$  and  $pQ_x = du/d\varphi$ , leads to

$$d^2u/d\varphi^2 + Q_x^2u = A_n Q_x u^2 \cos(n\varphi + \varphi_0) + \dots \quad (9)$$

where  $A_n = n^{th}$  harmonic component of  $SQ_x(\beta_x)^{5/2}$ .

Adjusting  $Q_x$  and  $S(s)$  transforms circular orbits in  $(u, \varphi)$  space into triangular separatrices, along which there is an exponential increase of  $x$  to the extraction septa, sited in dispersion free regions. Rf knock-out increases the emittance,  $\varepsilon$ , and is often used. Tune adjustments may be direct, or by rf steering (chromatic ejection). Other schemes use stochastic extraction or a high- $\beta$  insertion. For all momenta, separatrices need to be superimposed [8]. The unstable, fixed points are

at  $(r, \gamma)$  from the origin, given by  $r = 8((n/3) - Q_x)/A_n$ , and by  $\gamma = n\varphi/3 + \tan^{-1}(p/u) = (0, -2\pi/3, -4\pi/3) - \varphi_o/3$ .

## References

- [1] G.H. Rees, CERN 94-01 (1994) 731
- [2] C.R. Prior, G.H. Rees, NIM A415, No. 1, 2 (1998) 357
- [3] G.H. Rees, RAST, Vol. 1 (2008) 45
- [4] T.K. Kho, R.J. Lari, VIIIth Int. Conf. High Energy Acc., CERN (1971) 98
- [5] H. Takaki et al, PAC 07, 230
- [6] A. Franchi, S. Gilardoni, M. Giovannozzi, <http://ab-project-mte.web.cern.ch> & CERN Courier, 2009
- [7] V.M. Biryukov, V.I. Kotov, Yu.A. Chesnokov, Crystal Chanelling and its Applications at High Energy Accelerators, Springer, 1997
- [8] W. Hardt, CERN PS/DL/LEAR, Note 81-6

### 4.12.1 Aspects of Slow Extraction

P.J. Bryant, CERN

(See also Sec.4.12.) For irradiation of tumors in medical synchrotrons (Sec.1.6.14), the survival of the tissue surrounding the tumour depends not only on the spatial precision of the beam delivery, but also on the ability of the machine to deliver an exact radiation dose. The use of slow resonant extraction from a synchrotron [1] extends the beam spill time sufficiently to perform on-line dosimetry at the patient and to switch the beam on and off according to the dose required. Either the half-integer or third-integer resonance can be used, but the current trend is towards using the slower and more controllable spills from the third-integer resonance. Higher order resonances are not used because the angular separation of the separatrices becomes too small.

**Basic theory** A possible technique for designing an extraction system is to apply the *simplified Kobayashi Hamiltonian*,  $H$ , for the motion in the  $x$  plane [2, 3] and then to refine the design with tracking programs that include higher-order effects,

$$H = \frac{\varepsilon}{2}(X^2 + X'^2) + \frac{S}{4}(3XX'^2 - X^3) \quad (1)$$

where  $(X, X')$  are the normalised coordinates of a particle, the normalised sextupole strength is  $S = (1/2)\beta_{x,\text{sext}}^{3/2}(d^2B_y/dx^2)\ell_{\text{sext}}/(B\rho)$ ,  $\varepsilon = 6\pi\Delta\nu$ ,  $\delta\nu$  is the distance from the resonance  $\nu =$

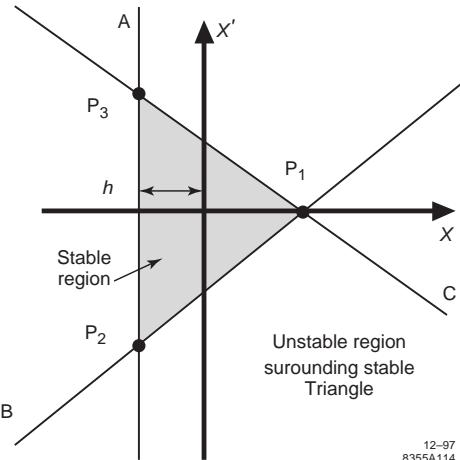


Figure 1: Phase-space geometry at the sextupole.  $h = 2\varepsilon/(3S) = 4\pi\delta\nu/S$ . Separatrix: A:  $X = -2\varepsilon/(3S)$ . B:  $-\sqrt{3}X'/2 + X/2 = 2\varepsilon/(3S)$ . C:  $\sqrt{3}X'/2 + X/2 = 2\varepsilon/(3S)$ . Fixed points: P<sub>1</sub>:  $(4\varepsilon/(3S), 0)$ . P<sub>2</sub>:  $(-2\varepsilon/(3S), -2\varepsilon/(\sqrt{3}S))$ . P<sub>3</sub>:  $(-2\varepsilon/(3S), 2\varepsilon/(\sqrt{3}S))$ .

$n \pm 1/3$ . Fig.1 summarises the phase-space geometry derived from Eq.(1) at the sextupole for mono-energetic particles whose equilibrium orbit is centred on the origin.

The sense of rotation in the stable region is clockwise when the particles are above resonance and the whole diagram will rotate clockwise if the observer moves downstream. The resonance sextupole is best put in a dispersion-free region. Elsewhere in the machine the dispersion will shift the equilibrium orbit and the stable triangle. The equation of a separatrix anywhere in the machine, including the effect of dispersion, is

$$(X - D_n\delta) \cos \alpha + (X' - D'_n\delta) \sin \alpha = h \quad (2)$$

where  $(D_n, D'_n)$  is the normalised dispersion function and  $\alpha$  is the angle measured from the  $X$ -axis in the counter-clockwise direction to the perpendicular,  $h$ , to the extraction separatrix. The spiral step,  $\Delta X_{\text{ES}}$ , at the electrostatic septum is

$$\Delta X_{\text{ES}} = \left[ \frac{4\varepsilon^2}{3S} - \frac{3S}{4}(X_{\text{ES}}^2 + X'_{\text{ES}}^2) \right] |\cos(\alpha - \frac{\pi}{2})| \quad (3)$$

**Hardt condition** When the beam enters the resonance, there is a range of momentum and emittance values that defines a boundary between the stable and unstable regions (Fig.2). The large amplitude/emittance particles will enter resonance

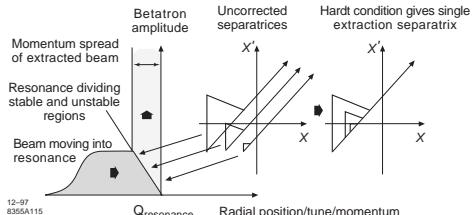


Figure 2: Hardt condition seen at electrostatic septum.

first, while still some distance from the resonant tune, and there is a continuous range of stable-triangle sizes down to zero for the zero-amplitude/emittance particles that must be exactly on resonance to be unstable. In general, the dispersion function will spread out this range of phase-space triangles and separatrices so the particles of different momenta will reach the electrostatic septum with different angles. The condition to set the optics such that all the extraction separatrices are aligned to give minimum losses on the electrostatic septum is known as the *Hardt condition* [4, 5],

$$D_n \cos \alpha + D'_n \sin \alpha = -4\pi\xi/S \quad (4)$$

where  $\xi = d\nu/(dp/p)$  is the chromaticity.

Effectively the Hardt condition moves the stable triangles along the direction of the dispersion vector by varying  $\xi$ . Fig.2 describes a general scenario, but the Hardt condition is only effective if the resonance is stationary and the beam is moved e.g. by a betatron core or stochastic rf noise.

Medical machines work below transition and ideally  $\xi < 0$  to ensure the transverse stability (Sec.2.4.9) of the “waiting” beam. Extraction is best made to the outside of the ring and injection to the inside. These choices limit the possibilities for applying the Hardt condition to positions where the electrostatic septum has  $D_n > 0, D'_n < 0$ , e.g. as in the second half of a dispersion bump.

**Configuration of the resonance** Fig.3 shows a possible layout for exciting the  $\nu_x = 1.666$  resonance. The lattice comprises two arcs with closed dispersion bumps linked by two dispersion-free insertions. The resonance sextupole is placed in one dispersion-free region. The electrostatic septum is placed in the second half of the arc where  $D > 0$  and  $D' < 0$  for the Hardt condition. The phase advance to the electrostatic septum must not be less than  $210^\circ$  or the next separatrix will be aperture limited on the septum. The phase advance to the magnetic septum can be as high as

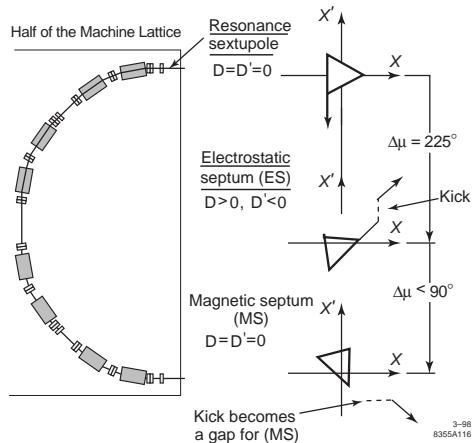


Figure 3: Possible layout of the extraction elements.

$105^\circ$  before the following separatrix hits the septum, but in practice an angle  $\sim 60^\circ$  is sufficient and more convenient.

**Sensitivity to ripple [6]** Irregularities of spill often quantified by *duty factor*  $F$ ,

$$F = \langle \phi \rangle^2 / \langle \phi^2 \rangle \quad \text{over time } T_{\text{spill}} \quad (5)$$

where  $\phi = dN/dt$ , the particle flux in the spill. The name duty factor is an unfortunate accident of history and should not be confused with duty cycle.

Typically, the transit time in the resonance varies from 50 to 4000 turns and most particles are extracted within 2000. For a revolution time of  $0.5 \mu\text{s}$ , this represents a delay of 1 ms. Thus, for common power supply, ripple frequencies below 1 kHz, it is reasonable to assume that the particle flux entering the resonance appears instantaneously in the spill. The particle flux in the spill is then directly proportional to the relative motion,  $\dot{\nu}$ , between the resonance and the waiting beam, i.e.  $\phi = \dot{N} = \lambda(\dot{\nu}_0 + \dot{\nu}_{\text{ripple}}) = \lambda(\dot{\nu}_0 + \omega\delta\nu \cos \omega t)$ . Substitution in Eq.(5) gives duty factor for ripple below 1 kHz,

$$F = \left[ 1 + \frac{1}{2} \frac{(\omega\delta\nu)^2}{\dot{\nu}_0^2} \right]^{-1} \quad (6)$$

The above assumes that  $\dot{\nu}_0 > \omega\delta\nu$ . The contrary is likely to be true and, once  $\dot{\nu}_0 \leq \omega\delta\nu$ , the resonance plunges periodically into the stack, chopping the spill into a series of spikes.

Above 1 kHz, the storage of particles in the resonance applies some smoothing and modifies

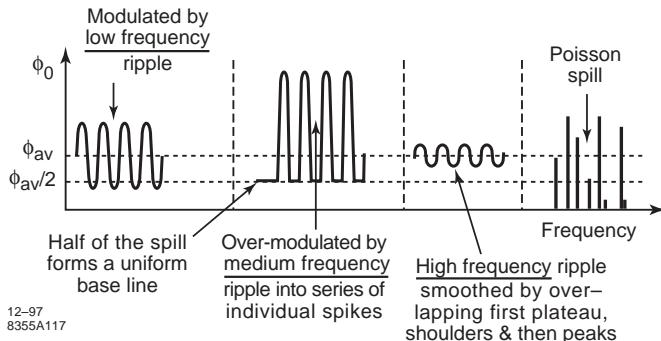


Figure 4: Typical characteristics of ripple at different frequencies.

Eq.(6) into (7). This occurs because the spikes have a characteristic shape with a narrow leading peak, a small shoulder and then a long plateau and the plateaux overlap [7]. Above 1 kHz, it is very likely that  $\nu_0 > \omega\delta\nu$  and the spill becomes a series of spikes. Duty Factor for ripple 1-100 kHz:

$$F = \left[ 1 + \frac{1}{4} \frac{(\omega\delta\nu)^2}{\nu_0^2} \right]^{-1} \quad (7)$$

Above 100 kHz, the leading spikes overlap and the spill is smoothed until the frequency is high enough to reveal the Poisson statistics in the beam density,

$$F = \bar{N}/(\bar{N} + 1) \quad (8)$$

where  $\bar{N}$  is the expected number of particles in a single measurement bin. Fig.4 summarises the characteristics of spills affected by low, medium, high and very high frequency ripple.

Eq.(6) shows that it is possible to reduce the influence of ripple by increasing the speed,  $\nu_0$ , at which the particles cross the boundary between the stable and unstable regions. This can be done by adding a band of stochastic noise to accelerate the beam [8], or by channelling the particles past an empty rf bucket [6].

**Emittance of the extracted beam** In a perfectly adjusted extraction scheme, the emittance  $\epsilon_x$  of the extracted beam will be extremely small. This can be seen from the Liouville theorem and the conservation of the combined longitudinal and transverse phase spaces,

$$\epsilon_{x,\text{stack}}\delta_{\text{stack}}T_0 = \epsilon_{x,\text{spill}}\delta_{\text{spill}}T_{\text{spill}} \quad (9)$$

The large ratio between the revolution time and the spill time ( $\sim 10^6$ ) is compensated by an equally large reduction in the transverse emittance

ratio. However, in measurements that are integrated over the spill, the true emittance may be masked by movements of the separatrix, changes in machine parameters, or dispersion effects. One effect that is not included in Eq.(9) but does give a true emittance in the extracted beam is the coupling from the vertical betatron motion in the excitation sextupole [9]. This effect can be estimated using

$$\frac{\epsilon_x}{\epsilon_y} \approx \frac{3\pi\sqrt{3}}{4} \frac{\beta_y}{\beta_x} S^2 (X_{\text{SE}}^2 + X'_{\text{SE}}^2) \sin 2\pi\nu_y t \quad (10)$$

where  $\epsilon_{x,y}$  are the geometric extracted emittances.

## References

- [1] H.G. Hereward, Proc. Vth Int. Conf. on High Energy Acc. (1963)
- [2] Y. Kobayashi, H. Takahashi, Proc. VIth Int. Conf. High Energy Acc. (1967) 347
- [3] M.Q. Barton, Proc. VIIIth Int. Conf. High Energy Acc. (1971) 85
- [4] W. Hardt, CERN/PS/DL/LEAR Note 81-6
- [5] C. Steinbach, EPAC 92, 857
- [6] R. Cappi, C. Steinbach, PAC 81
- [7] M. Pullia, CERN/PS 97-50 (DI)
- [8] S. van der Meer, CERN/PS/AA 78-6 (1978)
- [9] Private communication M. Pullia, TERA Foundation and CERN

## 4.13 COLLISION SCHEMES FOR RING COLLIDERS

*M. Zobov, INFN Frascati  
F. Zimmermann, CERN*

Present generation lepton factories have been very successful [1]. Both B-factories, KEKB in Japan

and PEP-II in USA, largely exceeded their design goals. The Italian  $\Phi$ -factory DAΦNE exceeded its phase-I design luminosity and further increased it by a factor 3 after implementing a novel crab waist collision scheme. The recently commissioned Tau-Charm factory in Beijing is well advanced towards its design luminosity. Similar progress is seen for hadron colliders, where, with much ingenuity, Tevatron and RHIC pushed their luminosities to new limits, and, only in its second year of operation (2011), the Large Hadron Collider (LHC) at CERN exceeded its design luminosity if scaled to the actual beam energy of 3.5 TeV [2] (Sec.4.1).

All the present generation ring colliders relied, at least in the beginning of their operation, on the standard strategy choice for high-luminosity parameters, which can be understood from expressions for luminosity  $L$  (Sec.4.1) and beam-beam parameter  $\xi_{x,y}$  (Sec.2.5.1). In the case of head-on collisions of identical short bunches they are

$$\begin{aligned} L &= n_b f_0 \frac{N_b^2}{4\pi\sigma_x^*\sigma_y^*} \\ &= n_b f_0 \frac{\pi\gamma^2 \xi_x \xi_y \epsilon_x}{r_e^2 \beta_y^*} \left(1 + \frac{\sigma_y^*}{\sigma_x^*}\right)^2 \end{aligned} \quad (1)$$

with

$$\xi_{x,y} = \frac{N_b r_e}{2\pi\gamma} \frac{\beta_{x,y}^*}{\sigma_{x,y}^*(\sigma_x^* + \sigma_y^*)} \quad (2)$$

where the asterisk indicates quantities at the interaction point (IP). This shows that the luminosity can be increased by:

1. raising number of particles per bunch,  $N_b$ ;
2. more colliding bunches,  $n_b$ ;
3. increase of the emittance,  $\epsilon_x$  and  $\epsilon_y$  (at constant  $\xi_{x,y}$ );
4. smaller beta functions at the IP,  $\beta_{x,y}^*$ ;
5. colliding beams with equal rms sizes at the IP,  $\sigma_y^* = \sigma_x^*$ ; and
6. pushing up the beam-beam parameters,  $\xi_{x,y}$ .

The present (and past) colliders obtained their good luminosity performances by trying to fulfill the above conditions as much as possible except that: A) for lepton factories transversely flat bunches are collided, with  $\sigma_y^* \ll \sigma_x^*$ , since the transverse emittances are naturally unequal and it is rather difficult to provide a good dynamic aperture in case the IP beta functions are very low in both transverse planes; for hadron colliders the beams are indeed round; B) in order to eliminate parasitic collisions in multibunch operation a non-zero horizontal crossing angle  $\theta$  has been necessary. In the factories a relatively small Piwinski

angle  $\phi = \sigma_z/\sigma_x^* \tan(\theta/2)$  avoided excessive geometric luminosity reduction and diminished the strength of synchrobetatron resonances excited by beam-beam interaction with a crossing angle.

However, further substantial luminosity increase based on the standard collision scheme is hardly possible due to several limitations imposed by beam dynamics requirements: A) In order to minimize the luminosity reduction due to the hour-glass effect (the variation of the vertical beam size within the longitudinal position along the collision region) the vertical beta function at the IP cannot be much smaller than the bunch length. B) A drastic bunch length reduction is impossible without incurring single bunch instabilities: bunch lengthening and microwave instabilities due to the beam interaction with the surrounding vacuum chamber. Besides, too short bunches tend to produce coherent synchrotron radiation (CSR) affecting beam quality and leading to a dramatic increase of the power losses. C) A multibunch current increase would result in different kinds of (coupled-bunch) beam instabilities, in excessive power loss due to interactions with parasitic higher order modes (HOM), resistive wall or due to electron cloud (especially for SC hadron rings), in an increase of the required wall plug power, or in unrealistic rf system parameters; D) Higher emittances conflict with stay-clear and dynamic aperture limitations, and require again higher currents to exploit the emittance increase for the luminosity enhancement. F) Tune shifts saturate and beam lifetime drops due to a strong nonlinear beam-beam interaction.

In order to overcome these (and other) limitations several novel collisions concepts and new collision schemes were proposed. The most popular approaches are the following:

1. round beam collision preserving additional integral of motion (e.g. [3]);
2. crab crossing [4, 5, 6];
3. collision with large Piwinski angle [7] (including “superbunch” schemes in hadron colliders [8, 9]);
4. collisions with alternating planes of crossing in different IPs [10];
5. (other) beam-beam compensation schemes (Sec.4.14);
6. colliding hadron bunches with longitudinally flat profile [9, 11, 12];
7. crab waist collision [13, 14];
8. collision with traveling waist [15];
9. longitudinal strong rf focusing [16];
10. collisions with negative momentum compaction factor [17], e.g. for shortening the bunches and mitigating longitudinal beam-beam effects; and

11. collisions at much higher beam energy with extremely large crossing angle [18, 19], e.g. for reducing the beam-beam tune shift, suppressing instabilities, and (for leptons) enhancing radiation damping.

The idea of round-beam collision was proposed more than 20 years ago for the Novosibirsk  $\Phi$ -factory design. It requires equal emittances, equal and small fractional tunes, equal beta functions at the IP, no betatron coupling in the arcs. A  $90^\circ$  rotation at each passage of the transverse oscillation plane by means of solenoids in the interaction regions (IR) provides conservation of the longitudinal component of the angular momentum  $M_z = yp_x - xp_y$ . Thus the transverse motion becomes one-dimensional. In addition to the obvious advantages from Eq.(1), the round beam scheme helps to eliminate all betatron coupling resonances that are of crucial importance for tune shift saturation and lifetime degradation. The synchrobetatron resonances are also weakened since the transverse tune shift is almost independent of the particle's longitudinal position along the bunch. The round beam concept was successfully tested at the electron-positron collider VEPP2000 in 2007 at 510 MeV beam energy [20]. Despite the low energy a high single bunch luminosity of  $10^{31} \text{ cm}^{-2}\text{s}^{-1}$  was achieved together with a maximum tune shift as high as 0.1. Another round beam collisions scheme, “Möbius accelerator” [21] was tested at CESR. It provided a tune shift of 0.09 in agreement with simulations [22].

In 1988 R. Palmer proposed the crab-crossing collision scheme for linear colliders [4]. Soon it was adopted and further developed for storage ring colliders [5]. This idea makes it possible to collide bunches at a large crossing angle without luminosity loss and without excitation of synchrobetatron resonances. In the crab crossing scheme both bunches are tilted prior to the collision, by half the crossing angle  $\theta/2$ , effectively providing a head-on collision at the IP. The tilt is created by a transverse rf deflector (crab cavity) imparting opposite transverse kicks to the bunch head and tail. The rf deflector is placed at a point where the betatron phase in the crossing plane is  $-\pi/2$  from the IP. In the classic crab crossing scheme another rf deflector after the collision point is used to restore the tilt. Crab crossing collision, with a single crab cavity per ring, was successfully implemented at the

KEK B-factory [24], where a world record luminosity of  $2.1 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1}$  was obtained in this configuration. However, the achieved luminosity remained lower than predicted by numerical simulation.

Colliding with a large Piwinski angle had been discouraged after the experience of harmful synchrobetatron resonances with 24-mrad crossing angle at the double storage ring DORIS-I [23], until, in 1995, Hirata suggested that a large angle might have several merits for future lepton factories [7]. Also for hadron colliders increases in the bunch length or the crossing angle have been proposed for luminosity optimization at the beam-beam limit [8, 9]. The advantages of a large  $\Phi$  can be understood by writing down the formulae for the luminosity and tune shifts with a horizontal crossing angle. Neglecting the hour-glass effect, these expressions can be obtained from Eq.(1) simply by substituting the horizontal beam size  $\sigma_x^*$  by the effective transverse size  $\sigma_x^*(1 + \Phi^2)^{1/2}$ . Then, for large Piwinski angle,  $\Phi \gg 1$ , the luminosity and the tune shifts scale as [25].

$$L \propto \frac{N_b \xi_y}{\beta_y^*}; \xi_y \propto \frac{N_b \sqrt{\beta_y^*/\epsilon_y}}{\sigma_z \theta}; \xi_x \propto \frac{N_b}{(\sigma_z \theta)^2} \quad (3)$$

If in such a case  $N_b$  is increased proportionally to  $\sigma_z \theta$ , the vertical tune shift  $\xi_y$  remains constant, and the horizontal tune shift drops like  $1/(\sigma_z \theta)$ , while the luminosity grows as  $\sigma_z \theta$ . In hadron colliders with round beams and alternating crossing at two IPs the scaling of both  $\xi_x$  and  $\xi_y$  is identical to the one for  $\xi_y$  in the above formulae [8, 9].

In the case of hadron bunches colliding at two IPs with alternating crossing at a large Piwinski angle, for the same bunch charge and the same beam-beam tune shift the luminosity of a uniform (or flat) longitudinal distribution is exactly  $\sqrt{2}$  times higher than for a Gaussian bunch profile [11]. Flat bunches could be generated by using higher harmonic rf systems [12], by accelerating and/or generating one or several long superbunches [8, 9] using a barrier bucket rf system (Sec.4.10.5), by introducing empty phase space in the centre of a coasting beam before bunching [26], by recombining with an empty rf bucket [27], or by redistributing surfaces in phase space [27].

The idea of using a “travelling” waist (focus) to compensate the luminosity reduction due to the hour-glass effect in circular colliders came from

linear colliders [28]. In the travelling waist collision scheme, the optical focal point depends on the longitudinal position of a particle within the bunch. In other words, particles with different longitudinal coordinates along the bunch in collision “see” the same and minimal beta functions when they encounter the center of the opposing beam. In circular colliders the travelling waist can be realized by a combination of accelerator components that provides a transformation described by a Hamiltonian  $H = H_0 - (zp_y^2)/2$  relating the longitudinal position  $z$  and the vertical momentum  $p_y$ . For example, the travelling waist with the crab crossing can be obtained by combining crab cavities and sextupole magnets [15].

The longitudinal strong rf focusing is an alternative way to obtain short bunches at the IP [16]. It consists in realizing a large momentum compaction of the lattice together with a strong rf gradient. In this regime the bunch length is no longer constant, but it is modulated along the ring [29] and can be minimized at the IP. In turn, if the main impedance generating elements of the ring are located where the bunch remains long, it is possible to minimize the strength of wake fields. This helps avoiding microwave instabilities and excessive bunch lengthening due to the potential well distortion. This concept was proposed as one possible option for the DAΦNE upgrade [30].

Contrary to the conventional strategy, the crab waist collision scheme requires small emittance, large Piwinski angle and larger crossing angle. There is no need to decrease the bunch length and to push beam currents beyond the values already achieved in the present factories. This scheme can greatly enhance the luminosity of a collider since it combines several potentially advantageous ideas: collisions with a large Piwinski angle, micro-beta insertions and suppression of beam-beam resonances using the dedicated “crab waist” sextupoles. For details of beam dynamics in crab-waist collision see Sec.2.5.2.

The crab-waist collision scheme has been successfully tested at the electron-positron collider  $\Phi$ -factory DAΦNE providing a 3-fold luminosity increase [31], in good agreement with numerical simulations. Fig.1 demonstrates the effect of the crab sextupoles on beam-beam blow up and beam tails in DAΦNE.

Several new collider projects seek to exploit the potential of the crab waist collision scheme. In particular, physics and accelerator communities

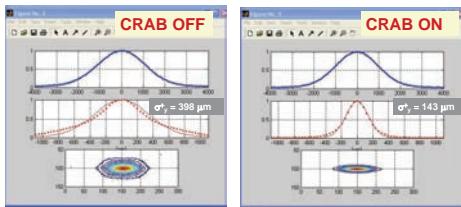


Figure 1: Transverse beam profiles in DAΦNE with crab-waist on and off.

are discussing Super B-factories (SuperB in Italy [32] and SuperKEKB in Japan [33]) and a Super tau-charm factory (Novosibirsk, Russia [34]) with luminosities about two orders of magnitude beyond those achieved at present or past B and tau-charm factories, respectively, an  $e^+e^-$  Higgs factory in the LHC tunnel [35], a 500-GeV  $e^+e^-$  collider in a 233-km tunnel [36], and an upgrade of the LHC, called HL-LHC, with a “virtual” peak luminosity of  $2 \times 10^{35} \text{ cm}^{-2}\text{s}^{-1}$  [37], all of which might make use of the crab-waist collision scheme, or, at least, incorporate some of its components.

## References

- [1] ICFA BDN 48, April 2009
- [2] <https://cern.ch/lpc>
- [3] V.V. Danilov et al, EPAC 96, 1149
- [4] R.B. Palmer, SLAC-PUB-4707, 1988
- [5] K. Oide, K. Yokoya, PR A40 (1989) 315
- [6] R. Calaga, R. Tomas, F. Zimmermann, CERN-2007-002, 77
- [7] K. Hirata, PRL 74 (1995) 2228
- [8] K. Takayama et al, PRL 88, 144801 (2002)
- [9] F. Ruggiero, F. Zimmermann, PRST-AB 5, 061001 (2002)
- [10] D. Neuffer, S. Peggs, SSC-063 (1986)
- [11] F. Ruggiero et al, KEK Proc. 2002-30, p. 131
- [12] C. Bhat, F. Zimmermann, IPAC 11, 1879
- [13] P. Raimondi, in 2nd SuperB Workshop (2006)
- [14] P. Raimondi et al, physics/0702033 (2007)
- [15] Y. Ohnishi, ICFA BDN 48 (2009) 252
- [16] A. Gallo et al, physics/0404020 (2004)
- [17] M. Zobov et al, physics/0607036 (2006)
- [18] P. Raimondi, DAΦNE upgrade workshop Alghero Sardinia (2003)
- [19] F. Ruggiero, M. Zobov, physics/0401129 (2004)
- [20] D. Berkaev et al, RuPAC 08 (2008) 70
- [21] R. Talman, PRL 74 (1995) 1590
- [22] E. Young et al, PAC 97, 1542
- [23] A. Piwinski, 11th HEACC Geneva (1980) 638

- [24] Y. Funakoshi et al, PAC 09, 2588
- [25] D. Shatilov, M. Zobov, ICFA BDN 37 (2005) p.99
- [26] J.P. Delahaye et al, 11th HEACC Geneva (1980) 299
- [27] M.Chanel, C. Carli, HB 2002 Chicago (2002)
- [28] V. Balakin, in SLAC-405 (1992)
- [29] A. Piwinski, NIM 72 (1969) 79
- [30] A. Gallo et al, EPAC 04, 683
- [31] M. Zobov et al, PRL 104, 174801 (2010)
- [32] M. Bona et al, arXiv: 0709.0451
- [33] K. Oide, Prog.Theor.Phys.122 (2009) 69
- [34] E. Levichev, Phys.Part.Nucl.L.5 (2008) 554
- [35] A. Blondel, F. Zimmermann, arXiv: 1112.2518 [hep-ex]
- [36] G.T. Lyons III, PhD thesis, arXiv: 1112.1105v1 [physics.acc-ph]
- [37] F. Zimmermann, CERN-2011-005 (2011) 317

## 4.14 BEAM-BEAM COMPENSATION SCHEMES

*J.P. Koutchouk, CERN  
V. Shiltsev, FNAL*

Significant experimental advances on beam-beam compensation have been made in the past decade. The perturbative or disruptive effect of the beam-beam interactions (See Sec.2.5.4) is indeed a basic limitation to increasing the luminosity of colliders and is a strong incentive to devise and study compensation methods. With the advent of hadron colliders with a small bunch spacing, it becomes necessary to consider the compensation of both the head-on and the long-range beam-beam effects. By nature, the head-on beam-beam force derives from a Poissonian potential while the magnetic force of optical lenses is Laplacian, defeating attempts at correcting one by the other, at least exactly. The long-range beam-beam effect is however close to Laplacian for realistic beam-beam separations, opening new compensation possibilities.

### 4.14.1 Compensation of the Head-On Beam-Beam Effect

**Four-beam compensation** If four beams are made to collide at the same point, with, for each direction of propagation, one beam of particles and one beam of antiparticles of equal intensity and transverse beam sizes, there is no net electromagnetic beam-beam force. This concept giving

exact compensation with a potential of substantially improved performance [1] was experimented in DCI ( $e^+/e^-$  at 0.8 GeV). While in a three-beam weak-strong configuration, an increase by a factor of 5 of the beam-beam limit was observed, no improvement of performance was obtained in the four beam configuration [2]. Unexpected excitation of non-linear beam-beam resonances was noticed, as well as coherent signals. These observations seem in qualitative agreement with the prediction [3] that the coherent beam-beam limit is not improved by the four-beam system, due to the cancellation of the beam-beam driven Landau damping. This coherent limit is however expected at somewhat higher beam-beam parameter than observed [4].

A four beam compensation concept has been contemplated for  $e^+/e^-$  linear colliders but found to be plagued by plasma instabilities, which lead to significant charge separation and luminosity reduction even at very small initial bunch displacement errors [5, 6].

**Electron-lens compensation** The compensation of the beam-beam effect by an auxiliary beam is a variation of the above solution that allows a drastic simplification appropriate for high-energy colliders, however with some limitations. An auxiliary electron beam of low energy is prepared in a source, made to collide with the main beam in a strong solenoidal field and dumped after the interaction, suppressing the possibility of coherent coupling, suspected to have plagued the four-beam concept. The auxiliary beam shall have suitable charge/direction of propagation for compensation and the same transverse positions and sizes as the perturbing beam [7]. It should ideally be positioned at the interaction point. This is however not possible in practice and mitigations must be carried out in the compensation strategy. The first study and implementation of an electron lens (TEL) was done at the Tevatron (see Sec. 7.2.17, with a schematic view of the lens). Comprehensive tests [8] have led to the optimization of the transverse electron current density for linear corrections (constant electron density over the main beam extent and smooth tails) and have demonstrated successfully the compatibility of electron lenses with the operation of superconducting hadron colliders. It was shown that the noise of the electron current can be reduced to a level that does not cause emittance blow-up of the main beam. This complex instrument reached

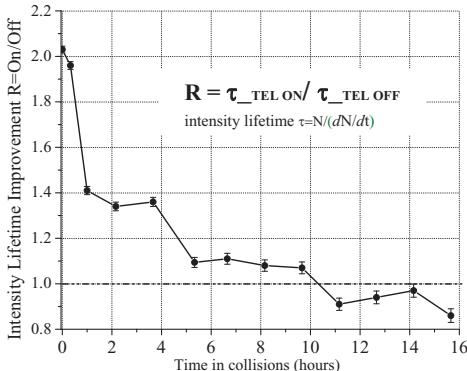


Figure 1: Improvement of the intensity lifetime of the Tevatron proton bunch 12 with TEL switched on, from [8].

a high reliability. It has been used to correct the primary beam-beam limitation of the Tevatron performance related to a tune spread along the bunch train. By using the lenses as pulsed bunch-by-bunch quadrupoles, the tunes of the normal and pacman bunches can be equalized, leading to a noticeable gain in lifetime (Fig.1). A full non-linear beam-beam correction has not yet been reported.

**Resonance compensation between interactions points** In colliders where the beams collide at several points, it is possible to phase these points such as to self-compensate certain resonant non-linear terms chosen or observed to be most harmful for the performance. For example [10], in the case of two identical collision points, all beam-beam resonances of order  $N$  can be canceled by choosing a betatron phase advance between collision points  $\Delta\Phi_x = \Delta\Phi_y = 2\pi(p/N + q)$ ,  $p, q, N$  integers with  $p$  odd. Detuning terms that are not phase-dependent will not be suppressed. The concept awaits experimental confirmation.

**Compensation by octupoles** In VEPP4 (e+/e- at 5.3 GeV), the compensation of the cubic beam-beam non-linearity using octupoles was systematically investigated by scanning the tune plane and the lattice octupole strength [9]. By measuring the beam loss rate, it was demonstrated that the width of resonant islands could be increased or decreased depending on the combination of beam-beam effect and powering of the lattice octupoles. In operating conditions, however, no

clear improvement could be obtained in VEPP-4, neither in CESR, where similar trials were made.

**Self-compensations in the crab-waist collision scheme** (Secs.2.5.2,4.13) One of the outstanding properties of the crab waist crossing scheme is the cancellation or large reduction of beam-beam resonances, thereby increasing the beam-beam limit.

#### 4.14.2 Compensation of the Long-Range Beam-Beam Effect

**Alternate crossing planes at pairs of collision points** It was recognized in [11] that crossings at pairs of identical interaction regions alternately in horizontal and vertical planes cancels the linear long-range beam-beam tune shifts. For sufficient beam separation where the perturbing beam can be well modeled by a wire, the magnetic field expansion is given by:

$$B_y + iB_x = \frac{\mu_0 I_b}{2\pi r_0} \sum_{n=1}^{\infty} (b_n + ia_n) \left( \frac{x+iy}{r_0} \right)^{n-1} \quad (1)$$

with  $(b_n + ia_n) = -\cos(n\Phi) - i\sin(n\Phi)$ ;

$I_b$  is the beam current,  $r_0$  the beam separation,  $\Phi = 0$  for horizontal separation and  $\Phi = \pi/2$  for vertical separation. One can note that the gradient coefficient  $b_2$  driving the linear tune shift changes sign from horizontal to vertical crossing. This is true for both normal and pacman bunches, that would otherwise experience different long-range beam-beam tune shifts. Hence alternate crossings effectively reduce the beam tune footprint. This cancellation depends on the multipole order and is not true in general. Alternate crossings excite beam-beam resonances in both planes, reducing the space available for the beam footprint in the tune plane. In simulations, e.g. [12], its benefit is not really determinable and experimental data are needed to assess the actual efficiency of this compensation method.

**Wire compensation** The principle of an almost exact compensation of the long-range beam-beam effect with wires was proposed in [13] for the LHC, where the long-range beam-beam interactions are clustered on either side of the interaction points. Each cluster is compensated by one wire positioned with the following requirements (Fig.1): small betatron phase advance between clustered perturbations and compensation (a few degrees); the same transverse beam aspect ratio as

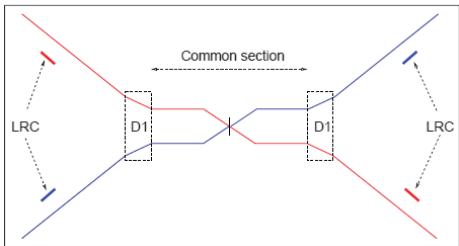


Figure 1: Schematic positions of wire compensators for a horizontal crossing. Trajectories of the two counter-propagating beams are shown in red and blue; “D1” represents the separation dipoles; and “LRC” the long-range compensators [13].

at the long-range perturbations; beam-wire separation identical to the beam-beam separation at the long-range interaction points when expressed in units of rms transverse size; integrated wire current equal to the sum of integrated beam currents in the cluster; sufficient separation between the beam channels to install the movable wire set-up.

Extensive numerical investigations and verifications of the compensation efficiency and robustness have been carried out, documented in the publications of the CARE-HHH, EuCARD-EuroLumi and USLARP workshops. Wire excitation and compensation units were fabricated, installed and studied in the SPS [14] and RHIC [15]. Compensation proper has been investigated in RHIC, where it resulted in detectable improvement of beam lifetime, and in the SPS, where one wire unit simulated the strong beam and another, installed at a betatron phase shift modeling the proposed wire set up for the LHC, compensated the perturbations induced by the former.

Figure 2 demonstrates the strength of the long-range beam-beam effect and the excellent compensation obtained in practice at the LHC tunes [16]. A residual tune dependence is reported, attributed to a non exact compensation related to some limitations in the SPS set-up.

If the long-range beam-beam interactions are not clustered but distributed all around the machine, wire compensation meets the same difficulties as non local non-linear corrections. It was simulated for the Tevatron and shown to yield a small improvement of stability if not a worsening [17]. On the contrary, for DAΦNE (INFN, Frascati), a non local wire correction was

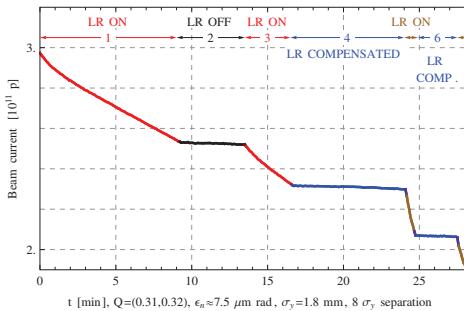


Figure 2: Beam current versus time in an LHC wire-compensator experiment at the CERN SPS. Exciting one wire models the effect of LHC long-range collisions. In addition exciting a second wire separated by a few degrees in betatron phase probes the efficiency of the compensation scheme.

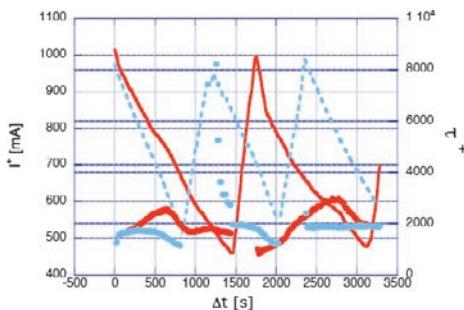


Figure 3: DAΦNE positron current and lifetime as a function of time: wires on (red) and wires off (cyan) [18].

successfully implemented, leading to a significant increase of the beam and luminosity lifetimes [18] (Fig.3).

**Compensation by multipoles** For non-local non-linear corrections, magnetic multipoles can be an alternative. A global approach proposed in [19] relies on a minimization of the nonlinearities of the one-turn map up to order 6 in the Hamiltonian, using magnetic multipoles in the machine arcs. It is shown, in a case close to that of the nominal LHC that the correction allows recovery of most of the 40% of dynamic aperture lost due to the beam-beam effect (long-range and head-on); see 4.

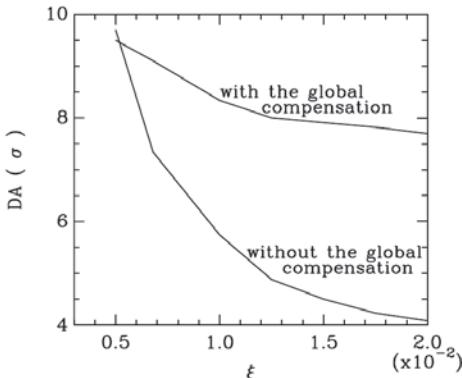


Figure 4: Dynamic aperture vs. beam-beam parameter without or with the global compensation of the long-range beam-beam interactions in the linear lattice [19].

## References

- [1] J. E. Augustin et al, Vol.2, Proc. 7<sup>th</sup> Int. Conf. HEA (1969) 113
- [2] Orsay Storage Ring Group, PAC 79, San Francisco
- [3] Ya. Derbenev, SLAC TRANS-151, 1973
- [4] B. Podobedov, R. Siemann, SLAC-PUB-6705, 1994
- [5] V.E. Balakyn, N.A. Solyak, Proc. 13<sup>th</sup> Int. Conf. HEA (1986) 151
- [6] D. Whittum, R. Siemann, PAC 97
- [7] E. Tsytanov et al., SSCL-519, (1993); Phys. Part. Nuclei 27 (1996) 279
- [8] W. Shiltsev et al., NJP 10 (2008) 043042
- [9] A.B. Temnykh, Proc. 3<sup>rd</sup> Adv. ICFA Beam Dynamics Workshop, Novosibirsk (1989)
- [10] S. Peggs, Proc. Workshop on AP Issues for SSC, UM HE 84-1 (1984) 58
- [11] D. Neuffer, S. Peggs, SSC-63 (1986)
- [12] F. Zimmermann, EPAC 04, 83
- [13] J.P. Koutchouk, LHC Project Note 223 (2000); PAC 01, 1681
- [14] F. Zimmermann et al, EPAC 04, 1936; PAC 05, 686
- [15] W. R. Calaga et al, PRST-AB 14, 091001 (2011)
- [16] e.g. in G. Sterbini, PhD thesis, EPFL, 2009, EuCARD Monographs Vol.6
- [17] T. Sen, B. Erdelyi, PAC 05, 2645
- [18] C. Milardi et al, EPAC 06, 2808
- [19] J. Shi, L. Jin, O. Kheawpum, PR E69, 036502 (2004)

## 4.15 SPACE CHARGE COMPENSATION (SCC) IN HADRON BEAMS

V. Shiltsev, FNAL

### 4.15.1 Longitudinal SCC: Inductive Inserts

Longitudinal space-charge fields can generate substantial distortion of the rf-generated potential wells, fill the extraction kicker gap in the beam, affect the incoherent synchrotron tune spread, and have the potential for causing instability and longitudinal emittance growth (also see Sec.2.4.5.1). The net effective voltage per turn resulting from the space-charge self voltage and the ring inductive wall impedance  $\omega_0 L$  is proportional to the slope of the beam current distribution  $e\beta c\lambda(s)$  and can be expressed as:

$$V_s = \frac{\partial \lambda(s)}{\partial s} \left[ \frac{g_0 Z_0}{2\beta\gamma^2} - \omega_0 L \right] e\beta c R$$

where  $R = c/\omega_0$  is the average machine radius,  $Z_0 = 377$  Ohm and  $g_0 = 1 + 2\ln(b/a)$  is the geometric space-charge constant,  $a$  and  $b$  are the beam radii and vacuum-chamber aperture. By introducing a tunable inductance  $L$ , e.g. ferrite rings, the term in brackets and, consequently, the space-charge effect may be substantially reduced or canceled at some chosen energy [1].

This concept has been experimentally proven at the LANL Proton Storage Ring at LANL where three inductive inserts, each consisting of 30 “cores” of a cylindrically shaped ferrite with thickness of 1 inch, inner diameter of 5 inches, and an outer diameter of 8 inches, were installed. The magnetic permeability of the ferrite could

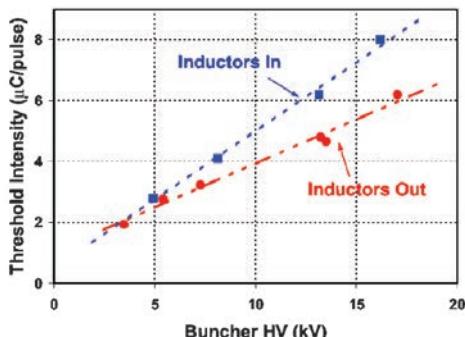


Figure 1: PSR e-p instability threshold vs rf buncher voltage with and without the inductive inserts.

be adjusted by introducing current into solenoids wound around the ferrite so that in the MHz range of frequencies the longitudinal space charge impedance of the machine was compensated [2]. A strong longitudinal instability was noticed at much higher frequencies of about 75 MHz, but it was later suppressed by heating the ferrite to a temperature of 130°C to make it more lossier.

The inserts have proven beneficial in raising the threshold for the two-stream e-p instability at PSR (Fig.1) and achieving shorter bunch length.

#### 4.15.2 Transverse SCC

There are several methods to compensate transverse s.-c. effects which often manifest themselves in the form of beam loss, core emittance growth, and halo formation (also see Sec.2.4.5.1).

In the case of **passive neutralization**, the s.-c. force of a proton beam can be compensated by ionization electrons, electron cloud, or negative ions which are approximately at rest longitudinally, but move transversely during the beam passage. The compensation condition reads  $\eta \approx 1/\gamma_p^2$ . Neutralized low energy beams of heavy ions have successfully been transported in a number of linear accelerators [3]. A factor of 9.5 increase of the maximum circulating beam current above (coherent) s.-c. limit was achieved at the Novosibirsk 1 MeV proton ring by increasing the residual gas pressure in excess of  $10^{-4}$  Torr and accumulation of ionization electrons [4]. The beam lifetime was very short and transverse and longitudinal proton distributions not well controlled.

Optimum compensation requires that the transverse electron and beam distributions are matched. That could be achieved by confining the electrons transversely with strong solenoid fields to “columns,” and using electrostatic electrodes to fine tune the charge density [5]. Strong magnetic field also stabilizes electron “column” motion and prevents coherent e-p instability. Simulations show significant reduction of s.-c. induced emittance growth with only few “columns” occupying a small fraction of the ring circumference [5].

**Electron lenses**, in which externally generated electron beam with matched transverse distribution collides with the proton beam inside a strong solenoid field, could also compensate the s.-c. tune shift [6]. Assuming the total length of the lenses  $L$ , distributed around the ring, and

an electron beams copropagating with the proton beam, the electron current needed per lens is [6]  $J_e = (B_f \kappa e c N_p / L) \beta_e / (\gamma_p^2 (1 - \beta_p \beta_e))$ , and for many accelerators of interest lies in the range of 1-10A for 10-40 keV electrons (here  $\kappa$  denotes the degree of compensation;  $B_f$  is proton bunching factor). These parameters are close to those of the operating Tevatron electron lenses (Sec.7.2.17). The SCC by electron lenses works better if the electron current is modulated to match longitudinal profile of proton bunches [7].

Passive cancellation of the next-to-leading term in the s.-c. force is possible by **octupole** fields. For a round beam, the 4th order of term of the direct s.-c. potential varies as  $(x^4 + 2x^2y^2 + y^4)$ , while the potential of an octupole is proportional to  $(x^4 - 6x^2y^2 + y^4)$ . Therefore, at least two families of octupoles are needed to reduce the s.-c. tune spread, which are placed at locations with either peak and intermediate values of the beta function, respectively. The beta functions should sufficiently vary over the length of an optical cell, e.g., by a factor 2 or more.

**Pole-face windings** allow precise adjustments of the tune shift with transverse position up to a high order. At the CERN ISR, 24 pole-face windings modifying the local magnetic field were used to correct the horizontal and vertical indirect s.-c. tune shift plus the next 4 orders in their Taylor expansions with respect to the horizontal position. The correction increased the maximum ISR beam current 15 times [8].

Recently proposed **fully nonlinear lattice** accelerators have promise to accommodate extraordinary large tune spreads in circulating beams without driving losses (resonance free optics) [9]. Numerical studies of s.-c. dynamics in such rings have been started.

#### References

- [1] R.J. Briggs, V.K. Neil, J.Nucl.Energy, Part C 8 (1966) 255
- [2] M.A. Plum et al, PRST-AB 2, 064201 (1999)
- [3] B.G. Logan et al, Nucl.Fusion, 45 (2005) 131
- [4] G. Dimov, V. Chupriyanov, PA 14 (1984) 155
- [5] V. Shiltsev et al, PAC 07, 1159; AIP 1086 (2009) 649
- [6] A. Burov, G. Foster, V. Shiltsev, FNAL-TM-2125 (2000)
- [7] M. Aiba et al, PAC 07, 3390
- [8] P.J. Bryant et al, CERN ISR-MA/75-54 (1975)
- [9] V. Danilov, S. Nagaitsev, PRST-AB 13, 084002 (2010)

## 4.16 OPERATIONAL CONSIDERATIONS ON COOLING

*S. Nagaitsyev, FNAL  
M. Blaskiewicz, BNL*

**Electron cooling** The electron cooling method is described in Sec.2.7.2. See also [1].

**Electron beam energy** At rather low voltages, it seems most natural to use direct electrostatic acceleration of electrons. Electron cooling devices of this type are operated in the range 500 V to 4.3 MV. At voltages higher than 200-300 kV, the high voltage generators and accelerating tubes are placed into special vessels filled with compressed SF<sub>6</sub> gas, which increases the electric breakdown field strength. At still higher voltages > 5 – 10 MV, one may have to consider electron linear accelerators in the ERL regime [2]. The electron energy ripple contributes to the effective longitudinal electron temperature. Energy regulation better than 100 ppm is commonly achieved.

**Electron beam current** Operating values of electron beam current are determined by the cooling rate required. A typical operating range of currents is from a few mA to ~1 A. With increase in the electron current the influence of space charge of the electron beam grows. A drift rotation of an electron beam occurs with respect to the axis, which is proportional to the electron density and inversely proportional to the longitudinal magnetic field. For weak magnetic fields, the electron drift velocity reaches the velocity spread of the cooled particles with relatively low electron current. As a rule, the cooling rate grows initially with increase of electron current and then it drops when the transverse velocities of the electrons become close to the ion transverse velocities.

Interaction of the electron and intense ion currents necessitated control of electron density in the beam cross-section. For this reason, special electron guns are developed, which allow for a noticeably decreased electron beam density at the center where the intense ion beam is stored [3]. Such a design enables the distribution of the cooling force values over the beam cross-section and optimizes the ion beam storage.

**Vacuum** It is well known that in devices employing electron beams one observes intense electron-induced gas desorption from the vacuum chamber surfaces. This kind of desorption is often one of the most important factors limiting the

performance and stability of such a device. In the beginning of electron beam operation, the vacuum deteriorates dramatically because of the desorption from the collector surface and one needs time to establish a maximum beam current. For copper (before conditioning), the desorption coefficient could be as high as 1-10 molecules per electron. Vacuum surface conditioning requires a specific electron beam dose of 1-2 mA-hour/cm<sup>2</sup> to reduce the desorption coefficient below 10<sup>-3</sup> [4]. After conditioning the pressure becomes a linear function of the electron beam losses outside of the collector. As a rule, these losses determine the vacuum level in the cooling section during routine operation.

Overall vacuum level in the storage ring must meet the requirements for obtaining an acceptable ion beam lifetime. Vacuum levels of 10<sup>-10</sup> – 10<sup>-11</sup> torr are commonly achieved in electron coolers.

**Secondary ions and electrons** On their way from the cathode to the collector the energetic primary electrons ionize the residual gas molecules. The ionized low energy ions can be trapped transversely by the electron beam potential and accumulated until it is neutralized. The trapped ions can have both positive (reduced drift velocities) and negative effects (electron beam instabilities) on cooling operations. Dedicated experiments to study electron beam neutralization were conducted at MOSOL [5] and LEAR [6]. The Fermilab 4.3-MeV electron cooler is particularly sensitive to ion beam accumulation because of its high energy and relatively weak focusing. Special measures were implemented to clear the ions from the electron beam [7]. Secondary (ionization) electrons may also be trapped under certain conditions and similarly to ions affect stability of the primary electron beam.

**Recombination** During the storage and cooling of highly charged ions, one of the main limiting factors in ion life time is ion recombination on the electron beam. Already in the first experiments at NAP-M, with magnetized electron cooling, it was experimentally demonstrated that the increase in the electron beam transverse temperature caused a weak decrease of the cooling rate but noticeably reduced recombination between protons and electrons (see also [8]). Special experiments have been carried out to verify the effect of reducing recombination by high electron temperature for the highly charged ions at GSI in the ESR storage ring [9].

**Transverse emittance of the ion beam** Cooling of intense ion beams is often limited by space charge effects. The mutual repulsion of ions inside the beam prevents further compression of the transverse emittance at a level determined by the value of the betatron tune shift due to space charge of the ion beam. As was learned from many experiments [10], with increasing beam intensity the tune shift rises up to  $\Delta\nu = 0.1 - 0.15$ , and from then on the emittance grows proportionally to the ion beam current, so that the beam density remains constant. It is of interest that in such a situation, the ion energy spread inside the beam grows very slowly with beam intensity.

**Ion beam intensity limits** Electron cooling is designed to decrease ion beam emittances and to extend its lifetime. The decrease in the energy spread shifts the ion beams closer to the threshold of coherent instabilities the development of which is accompanied by beam loss [11].

In most electron cooling rings these instabilities have been successfully suppressed by feedback systems. At several cooler rings ion beam losses not accompanied by clear coherent signals were observed. The effect is called “electron heating” and is described in [1].

**Stochastic cooling** (Sec.2.7.1) These systems involve wide-band feedback loops operating at high frequencies [12]. The operational considerations are broadly categorized as design, startup and stability issues.

**Design** During the design of a stochastic cooling system one needs to consider dynamic range [13], bandwidth, gain uniformity and delay stability issues. Dynamic range is particularly important for bunched beams, where the Schottky power can be less than 1% of the power in coherent lines [14]. Bandwidth, and the upper frequency in particular, is relevant for the total allowed delay. For longitudinal filter cooling one needs to consider the phase characteristics of the one turn delay as well [15]. The delays need to be held to better than a tenth of a wavelength for optimal cooling and better than that for a notch filter with bunched beams [16]. That implies feedback on the delays. In RHIC system transfer functions were taken periodically and these measurements used to adjust the gains and phases in the feedback loop.

The microwave link used for longitudinal cooling was stabilized with a pilot tone [17]. The detected phase variations were used to correct the

phase of the kicker signal by real time corrections to the I/Q modulators for each channel.

Another practical concern is the possibility of spurious coupling between the kicker and pickup via microwave transmission through the beam pipe. While these can often be found simply by running the system without beam we note that the RHIC multipath problem involved the kicker of one ring talking to a pickup in the other ring. This was solved by running the narrow band kickers in the two rings at different frequencies [18].

**Startup** During startup it is necessary to get initial setpoints for delays, gains and phases. Perhaps the most fundamental thing is to make sure that the delay is correct so that the cooling signal obtained from a certain set of particles is used to apply a kick to that same set. For wide band kickers the absolute time delay has to be correct to 1/10 of the wavelength across the entire band. Discrete frequency systems only require the phase to be correct at each frequency, independently. The coarse timing is easily accomplished by comparing the relative timing of the beam induced signal on the cavity to the cooling signal on the cavity. Once the initial timing is done the fine phase is adjusted by maximizing signal suppression. A fundamental quantity in this regard is the open loop system transfer function. One very important practical note is that the bunched beam transfer function can be well approximated by a coasting beam transfer function when there is significant synchrotron sideband overlap. Coasting beam transfer functions are well known [19] and the effect of signal suppression and filtering are straightforward to calculate. This is very helpful during startup since one can calculate beam transfer functions and the effects of signal suppression beforehand which helps when adjusting filters.

**Stability** It is necessary to maintain stable gains and delays for optimal cooling. In RHIC these changes are due to thermal effects in signal transmission and variations in resonant frequency as well as evolution of the beam momentum distribution. We employ slow feedback loops for correction. For longitudinal filter cooling, variations in the temperature of the optical fiber one turn delay are compensated by an optical trombone. We adjust the one turn delay by taking the system transfer function through the one turn delay filter. The onboard computer of the network analyzer orchestrates the various switches, measures the depth and location of the notch, and sends an

updated signal to computer controlled delay and attenuators, correcting the filter. This loop runs every few minutes and corrects for thermal drifts. Another correction loop inserts the network analyzer in the cooling loop. The gain and phase of the open loop system transfer function is then compared with ideal values and the amplitude and phase of the feedback path are adjusted to recreate the ideal values. This loop operates every few minutes and is controlled by the network analyzer.

## References

- [1] V.V. Parkhomchuk, A.N. Skrinsky, RAST 1 (2008) p.237
- [2] I. Ben-Zvi et al, PAC 05, 2741
- [3] A. Bubley et al, EPAC 02, 1356
- [4] A.N. Sharapa, A.V. Shemyakin, NIM A404 (1989) 185
- [5] A. Burov et al, Budker INP preprint 89-116; CERN/PS 93-03 (AR)
- [6] J. Bosser et al, NIM A391 (1997) 103
- [7] A. Shemyakin et al, IPAC 10, 867
- [8] H. Danared et al, PRL 72 (1994), 3775
- [9] P. Beller et al, NIM A532 (2004) 427
- [10] S. Nagaitsev et al, PAC 95, 2937
- [11] J. Bosser et al, NIM A441(2000) 1
- [12] J. Marriner, NIM A532 (2004) 11
- [13] R. Pasquinelli, AIP Conf 319 (1993) 58
- [14] D. Boussard, Lecture Notes in Physics, 296 (1988) 269
- [15] F. Caspers CERN-PS-2001-17 (RF) (2001)
- [16] G. Carron, L. Thorndahl, CERN-ISR-RF/78-12 (1978)
- [17] K. Mernick et al, BIW 10, 389
- [18] M. Blaskiewicz, J.M. Brennan, K. Mernick, PRL 105, 094801 (2010)
- [19] S. van der Meer, CERN/PS/AA/80-4 (1980)

# Chapter 5. MECHANICAL CONSIDERATIONS

## 5.1 MECHANICAL AND THERMAL PROPERTIES OF STRUCTURAL MATERIALS

*M. Kuchnir, FNAL*

The design of mechanical structures for superconducting accelerator components requires data on the typical mechanical properties and their temperature dependence as well as thermal conductivity and specific heat. The variety of materials used for this purpose is not very large: stainless steels, fiber reinforced epoxy composites, alloys of aluminum, copper, nickel and titanium.

An excellent source of information is the book by Reed and Clark [1]. Here we collected data from this and other references but we should keep in mind that most of the data presented here are typical properties and *not* design data.

In particular, data on strength of common metals and alloys could be presented in the form of bands instead of lines [2] since experimental values can vary as much as a factor of 2 or 3.

Good sources of data on thermal conductivity are Childs' [3], Touloukian's [4] and Campbell's [5] compilations. The collection by Touloukian includes books on other properties like specific heat and is not restricted to low temperatures. The Campbell loose leaves handbook stresses mechanical, thermal, electrical and magnetic properties of structural materials for superconducting machinery. Fig.1 includes data from these sources besides original measurements like the one in [6, 7] on composites of fiberglass and epoxy (G10-CR and G11-CR).

The data on specific heat (Fig.2) is largely based on the compilation [8]. For stainless steel, we used data from the original works of [9, 10] and for fiberglass epoxy from [11, 6].

The thermal expansion data (Fig.3) is presented in its more useful form, i.e. integrated and referenced to room temperature. Most of it

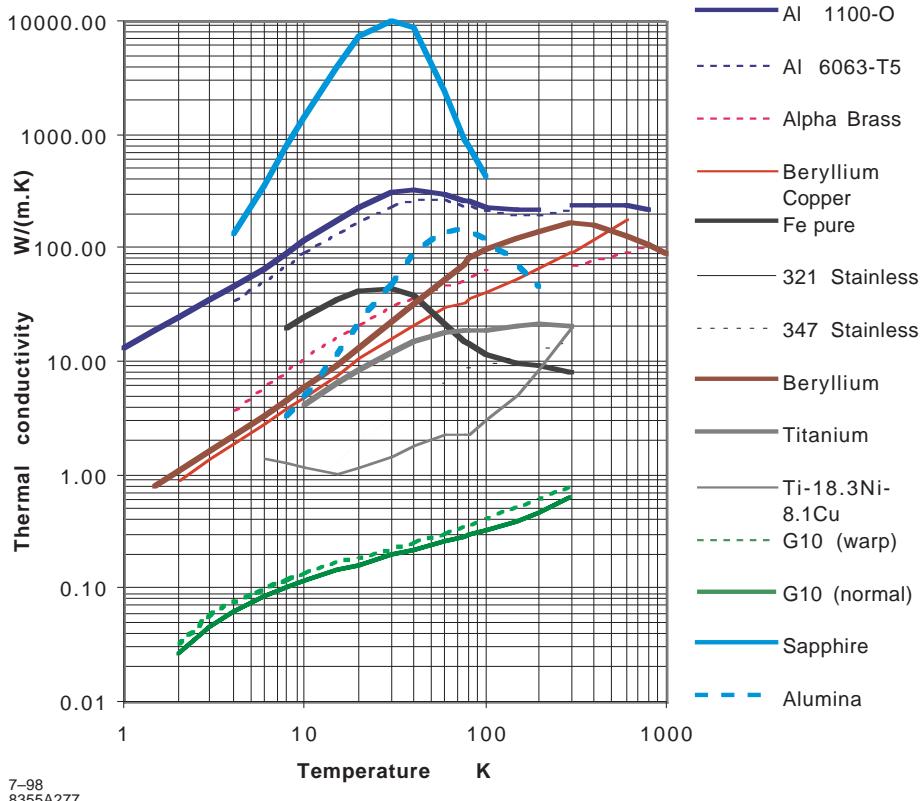
is based on the compilation [12] and the original work on composites [13].

Data on mechanical properties (Figs.4-7) were also obtained from secondary sources like the Metals Handbook [14], Smithells Metals Reference Book [15], Handbook of Stainless Steels [16], Marks' Standard Handbook of Mechanical Engineering [17], special topic books like Titanium: A Technical Guide [18], Aluminum [19], or original work like Ledbetter's [20] or Schwartzberg et al [21]. A very useful list of references of cryogenic properties of materials has been compiled by Weisend et al [22].

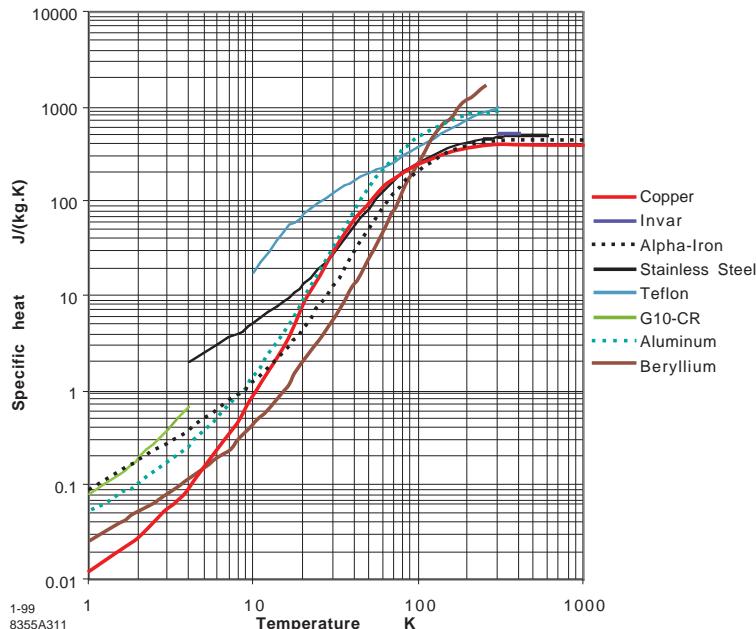
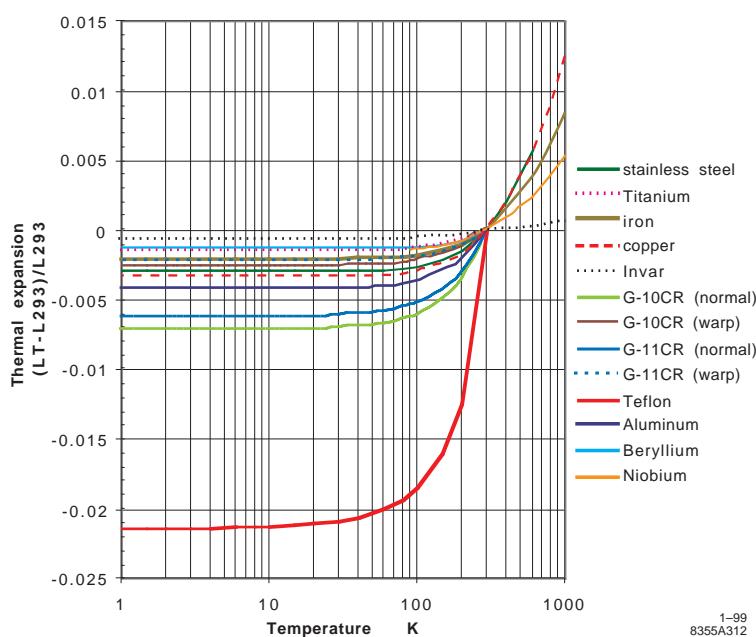
Additional information on properties of materials is available through the Internet. At the time of this printing some of these web sites are:  
<http://www.astm.org>  
<http://www.nist.gov/srd/intro.cfm>  
<http://www.matweb.com/>

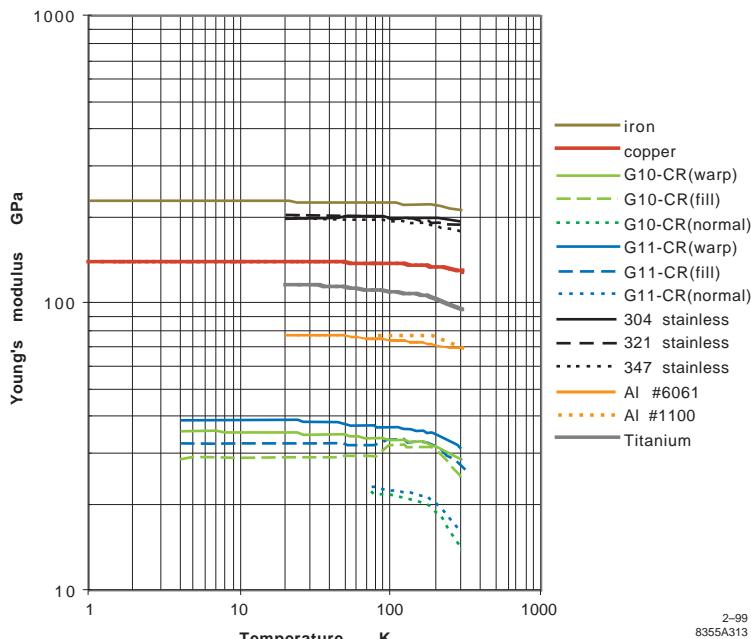
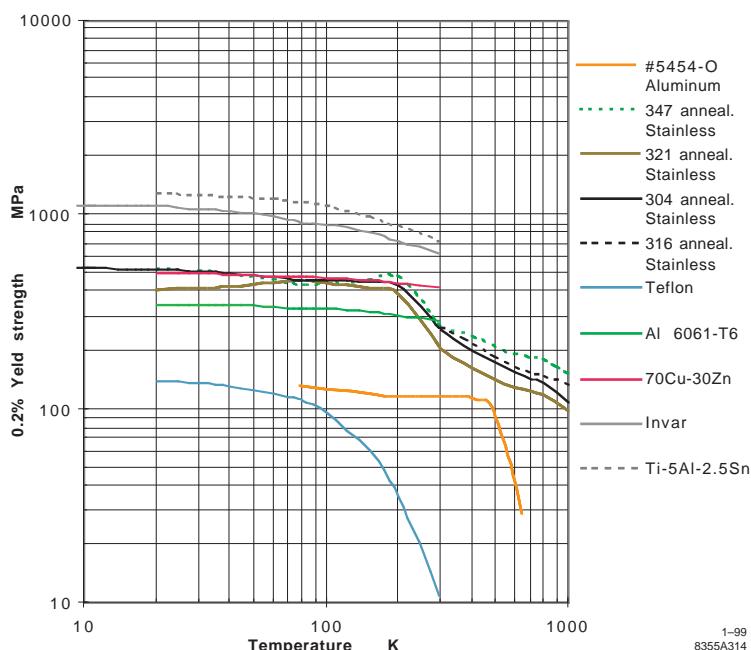
## References

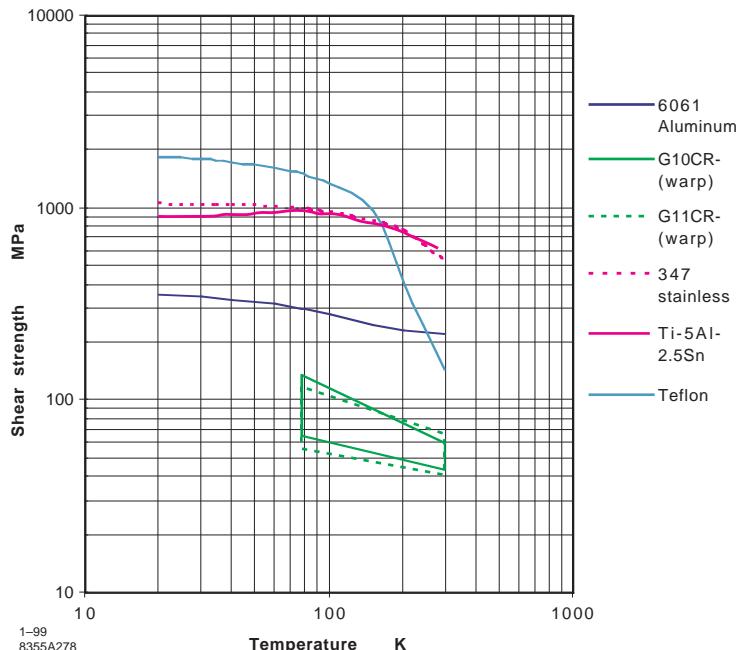
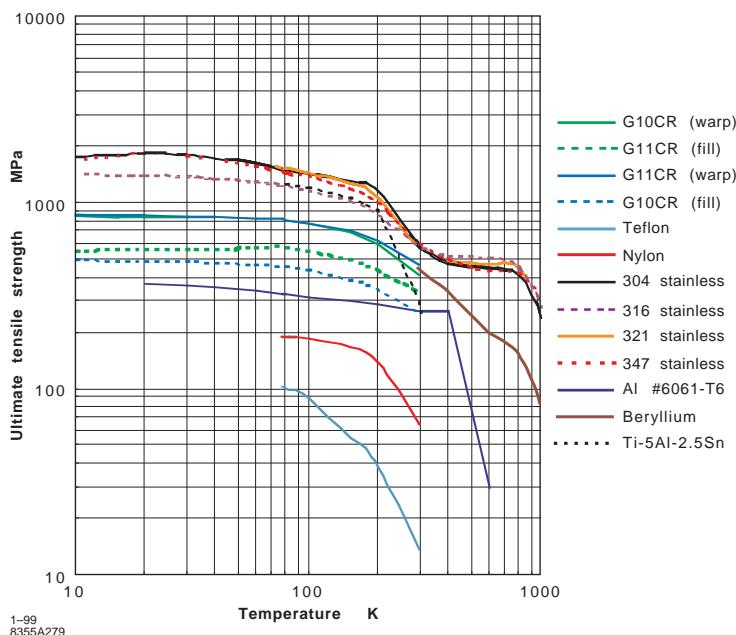
- [1] R.P. Reed, A.F. Clark, Materials at Low Temperatures, American Society for Metals, Metals Park, OH (1983)
- [2] R.B. Scott, Cryogenic Engineering, D.Van Nostrand (1959)
- [3] G.E. Childs, L.J. Ericks, R.L. Powell, Thermal Conductivities of Solids at Room Temperature and Below, NBS Monographs, National Bureau of Standards (1973) Vol.131
- [4] Y.S. Touloukian et al, Thermalconductivity - Metallic Elements and Alloys, Thermophysical Properties of Matter, IFI/Plenum(1970) Vol.1
- [5] J.E. Campbell, Batelle - Columbus Laboratory, Handbook on Materials for Superconducting Machinery (1976)
- [6] M.B. Kasen et al, Advances in Cryogenic Engineering-Materials 26, Plenum Press (1979) p.235
- [7] F.J. Walker, A.C. Anderson, Review of Scientific Instruments 52 (1981) 471

Figure 1: Thermal conductivity vs  $T$ .

- [8] R.J. Corruccini, J.J. Gniewek, National Bureau of Standards Monograph 21, Specific Heats and Enthalpies of Technical Solids at Low Temperatures (1960)
- [9] J.M. Corsan, I. Mitchem, 6th Int.Cryogenic Engineering Conf., IPC Science and Technology Press (1976) p.342
- [10] J.M. Corsan, N.I. Mitchem, Cryogenics (1979) 11
- [11] J.R. Benzinger, Advances in Cryogenic Engineering 26, Plenum Press (1979) p.252
- [12] R.J. Corruccini, J.J. Gniewek, National Bureau of Standards Monograph 29, Thermal Expansion of Technical Solids at Low Temperatures (1961)
- [13] M.A. Ranney, A.F. Clark, NBS/DoE Workshop on Materials at Low Temperature (1979)
- [14] H.E. Boyer, T.L. Gall, Metals Handbook, American Society for Metals (1985) Vol. Desk Edition
- [15] E.A. Brandes, Eds., Smithells Metals Reference Book, Butterworths
- [16] D. Peckner, I.M. Bernstein, Eds., Handbook of Stainless Steels, McGraw-Hill (1977)
- [17] T. Baumeister, E.A. Avallone, T.Baumeister. III,Eds., Marks' Standard Handbook for Mechanical Engineers, McGrawHill (1978)
- [18] M.J. Donachie, Eds., Titanium: A Technical-Guide, ASM International (1988)
- [19] J.E. Hatch, Eds., Aluminum, American Society for Metals
- [20] H.M. Ledbetter, Cryogenics (1980) 637
- [21] F.R. Schwartzberget al, Cryogenic Materials Data Handbook, Report ML-TDR-64-280 (1970)
- [22] J. G. Weisend II, V. Flynn, E. Thompson, R. P. Reed, A Reference Guide for Cryogenic Properties of Materials, SLAC-TN-03-023 (2003)

Figure 2: Specific heat vs  $T$ .Figure 3: Thermal expansion integrals vs  $T$ .

Figure 4: Elastic modulus vs  $T$ .Figure 5: 0.2% yield strength vs  $T$ .

Figure 6: Shear strength vs  $T$ .Figure 7: Ultimate tensile strength vs  $T$ .

## 5.2 MECHANICAL AND THERMAL PROPERTIES OF COMPOSITE SUPERCONDUCTORS

R.M. Scanlan, LBNL

**Composite conductors, mechanical** (See Sec.5.1 for elements and alloys.)

Modern superconductors are composites of the superconductor together with other elements that are fabricated with the superconductor and are an integral part of the composite. These elements usually include a high purity normal metal such as Cu, Ag, Al, as well as structural materials such as stainless steel. In the elastic range, the modulus can be estimated well from a simple rule of mixtures. For NbTi alloy based superconductors, the mechanical strength is usually dominated by the superconductor component which is heavily cold worked and precipitation hardened in order to optimize the superconducting critical current density. Example stress-strain relations [1] are shown in Fig.1. The modulus for different compositions can be estimated using the rule of mixtures, Eq.(1).

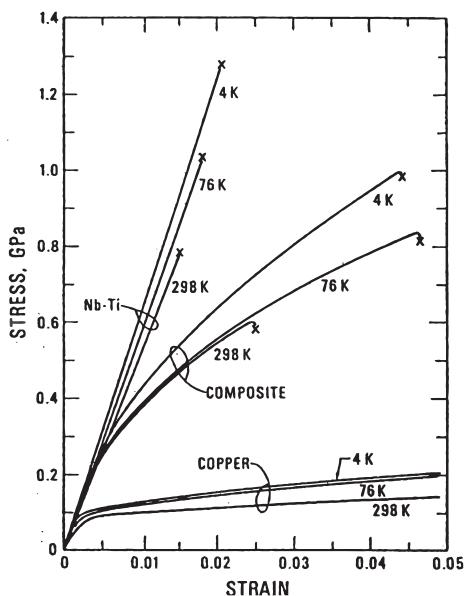


Figure 1: Stress-strain characteristic for a typical multifilamentary NbTi conductor and its components containing 36 vol.% NbTi and 64 vol.% Cu.

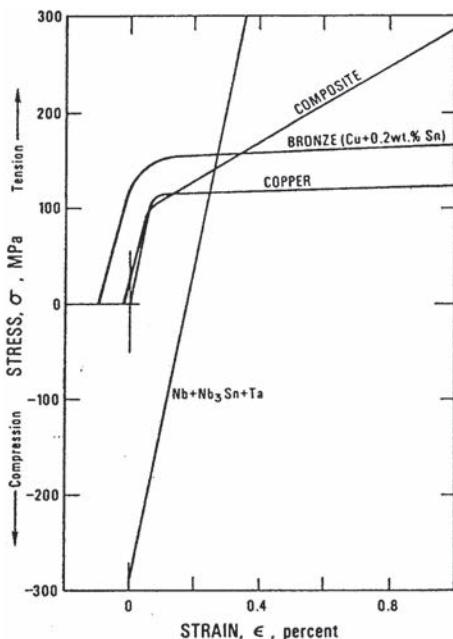


Figure 2: Stress-strain curves at 4K for a typical  $\text{Nb}_3\text{Sn}$  composite and its components, containing 11 vol.%  $\text{Nb}_3\text{Sn}$ , 1 vol.% unreacted Nb, 24 vol.% bronze, 2 vol.% Ta (diffusion barrier) and 62 vol.% Cu stabilizer.

For  $\text{Nb}_3\text{Sn}$  and other A15 superconductors the effective strain limit of the composite is determined by the brittle nature of the superconductor which fractures at about 0.2% tensile strain. The useful strain limit of the composite conductor is extended by the matrix which places the  $\text{Nb}_3\text{Sn}$  in compression due to the differential thermal expansion. The overall strain state [2] can be estimated from the rule of mixtures as is illustrated in Fig.2 using the component thermal expansion properties.

The HTS superconductors Bi-2223, Bi-2212, and YBCO are all brittle materials analogous to  $\text{Nb}_3\text{Sn}$ . They can withstand a limited strain range where the change in critical current is reversible, i.e. it recovers when the strain is removed. Above this strain limit, the critical current is permanently reduced. YBCO is fabricated on thick substrates, either Hastalloy or nickel alloy. These substrates greatly improve the mechanical properties of the composite conductors so that typical values [3] are 122 GPa (modulus) and 1000 MPa (tensile

yield strength at 77 K). The mechanical properties of Bi-2212 are less well-defined [4] and the safe operating range is still being explored.

Using the NbTi twisted filament and oxygen free Cu data one can predict their composite Young's modulus ( $E$ ) knowing the volume ratio (1.8 part Cu,  $X_{\text{Cu}}$ , to 1 part NbTi,  $X_{\text{NbTi}}$ ) using the rule of mixtures relationship,

$$E_{\text{comp}} = \frac{X_{\text{Cu}} E_{\text{Cu}} + X_{\text{NbTi}} E_{\text{NbTi}}}{X_{\text{Cu}} + X_{\text{NbTi}}} \quad (1)$$

**Composite conductors, thermal** Thermal conductivity values vary over several orders of magnitude for the different materials involved in SC magnet windings. They also vary by over an order of magnitude for the individual materials over the temperature range from 300 K to 5 K (see Fig.1, Sec.5.1). Thermal conductivity of pure metals can vary significantly depending on the state of cold work and of purity. Fig.3 [5] shows both effects for copper.

Fig.4 [6] shows the purity effect for Nb while Fig.5 [9] shows the influence of magnetic field on Nb.

Given the wide variation of thermal conductivity one needs to know the state of cold work and purity before making calculations over extended temp. ranges. Direct measurement of the as delivered material is best.

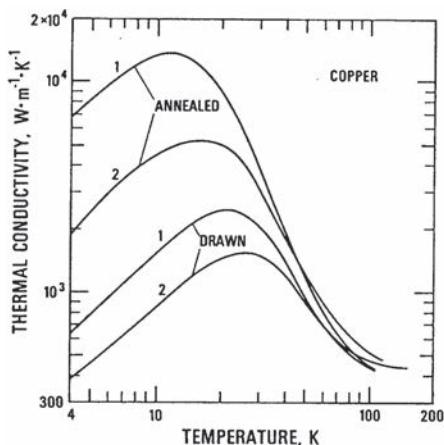


Figure 3: Thermal conductivity changes due to cold work and purity variation [5].

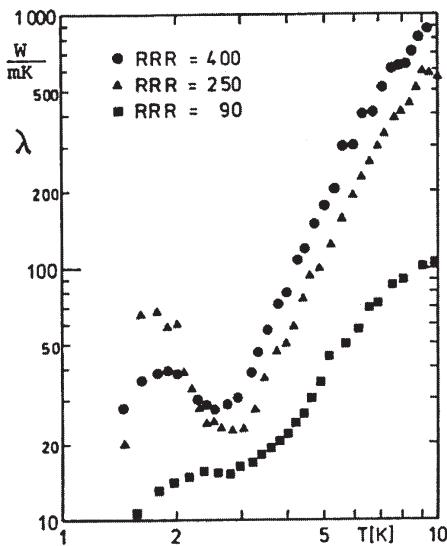


Figure 4: Thermal conductivity of Nb [6] for various purity levels. (RRR = residual resistivity ratio)

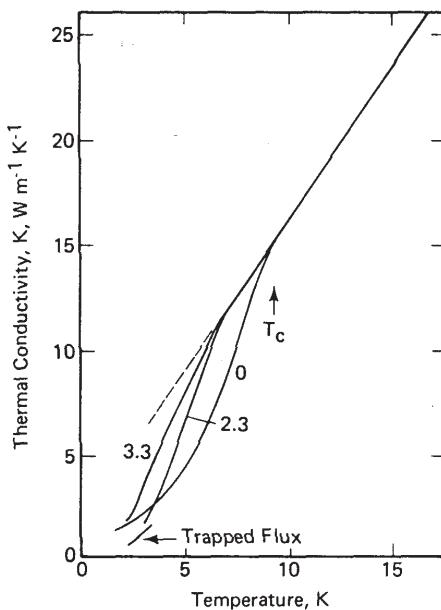


Figure 5: Influence of magnetic field on thermal conductivity of Nb [9]. Parameter on curve is magnetic field in Tesla.

Table 1- contraction of composites

| Coil                         | $\Delta T$ (K) | %contraction |
|------------------------------|----------------|--------------|
| NbTi/Kapton                  |                |              |
| SSC-inner, longitudinal [5]  | 294-4          | 0.26         |
| radial                       | 294-4          | 0.42         |
| Nb,Sn/ fiber-glass-epoxy [6] |                |              |
| outer                        | 293-77         | 0.26         |
| inner                        | 293-77         | 0.29         |

Thermal expansion coefficients likewise can vary significantly between materials and as a function of temperature (see Fig.3, Sec.5.1). Several measurements for composite coil sections have been reported and are shown in Tab.1. The rule of mixtures usually gives good results provided that the composite does not delaminate.

## References

- [1] R.P. Reed, R.P. Mikesell, Adv. in Cryogenics, V.22, Plenum (1977) p.463
- [2] D.S. Easton et al, J. Appl. Phys. 51 (1980) 2748ff
- [3] Ekin J W 2006 Experimental Techniques for Low-Temperature Measurements (Oxford: Oxford University Press)
- [4] D.R.Diederich et al, IEEE Trans. on Applied Superconductivity 11, p.3577, March 2001
- [5] R.P. Reed, A.F. Clark, Materials at Low Temperatures, American Society for Metals (1983)
- [6] Courtesy Wuppertal University
- [7] Courtesy C. Goodzeit, Brookhaven Nat. Lab.
- [8] D. Dell'Orco, H Tsui, Mechanical Measurements on D20 Nb<sub>3</sub>Sn Cables SC-MAG (1994) 465
- [9] E.W. Collings, Applied Superconductivity, Metallurgy and Physics of Titanium Alloys, V.1, Plenum Press (1986) p.260

## 5.3 THERMODYNAMIC & HYDRODYNAMIC PROPERTIES OF COOLANTS & CRYOGENS

*M. McAshan, FNAL*

**Critical and triple-point properties and thermodynamic states** With temperature and pressure as independent variables, the familiar relationships between the solid, liquid and gas phases of a pure substance appear as in Fig.1. Points on such a diagram are termed thermodynamic states, and in general, the laws of thermodynamics speak to the stationary relationships or equilibria that exist between such states.

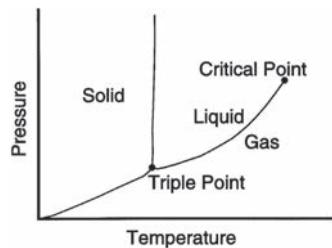


Figure 1: Schematic  $P$ - $T$  diagram for a pure substance.

The gas and solid phases of the substance are in equilibrium along the sublimation line, the solid and liquid along the melting line, and the liquid and gas along the saturation line. The intersection of the three lines marks a point of equilibrium of all three phases, the triple point, and the saturation line extends from the triple point to end in the critical point.

The liquid phase, therefore exists only between the triple point and the critical point. A line of constant pressure drawn through this region intersects both the melting line and the saturation line. These points of intersection occur at the freezing point temperature and the boiling point temperature of the liquid at the particular pressure. The normal freezing and boiling points are those at the pressure of 101.3 kPa (1 atmosphere).

At temperatures greater than critical, the fluid is said to be in the gas phase. At temperatures less than critical and pressures less than saturation, the fluid is said to be unsaturated or superheated vapor. If the temperature is less than critical and the pressure is greater than saturation, the fluid is unsaturated or subcooled liquid. The term supercritical fluid is somewhat ambiguous; it can be taken to mean a fluid at conditions greater than critical of either temperature or pressure.

Triple and critical point data for a number of commonly used substances are listed in Tab.1. Normal boiling points are given in Tab.2. Note that the triple-point pressure of carbon dioxide is greater than 1 atmosphere. Thus this substance has no normal boiling conditions, and instead, the saturated solid, dry ice, is in equilibrium with the saturated vapor at 1 atmosphere. This situation is also found in sulfur hexafluoride, and Tab.2 shows in this case that at 20°C (293.15 K) the saturation pressure is 2.103 MPa.

Table 1 Critical Properties of Fluids [4,8,13,19]

|                     |                                | Molwt  | Critical Point |                 |                    | Triple Point        |                 |
|---------------------|--------------------------------|--------|----------------|-----------------|--------------------|---------------------|-----------------|
|                     |                                |        | Temp.<br>K     | Pressure<br>MPa | Density<br>kg/cu-m | Temp.<br>K          | Pressure<br>kPa |
| Helium              | He                             | 4.0026 | 5.1953         | 0.22746         | 69.64              | <sup>a</sup> 2.1768 | 5.0418          |
| Hydrogen (para)     | H <sub>2</sub>                 | 2.0159 | 32.938         | 1.2838          | 31.36              | 13.81               | 7.042           |
| Neon                | Ne                             | 20.18  | 44.4           | 2.71            | 483.0              | 24.54               | 43.19           |
| Nitrogen            | N <sub>2</sub>                 | 28.01  | 126.2          | 3.399           | 314.03             | 63.15               | 12.54           |
| Argon               | A                              | 39.95  | 150.7          | 4.865           | 513.0              | 83.80               | 68.71           |
| Oxygen              | O <sub>2</sub>                 | 32.00  | 154.576        | 5.043           | 436.15             | 54.34               | 0.1453          |
| Methane             | CH <sub>4</sub>                | 16.04  | 190.7          | 4.599           | 160.43             | 90.68               | 11.74           |
| Krypton             | Kr                             | 83.80  | 209.40         | 5.492           | 908.51             | 115.77              | 73.154          |
| R-14                | CF <sub>4</sub>                | 88.01  | 227.5          | 3.745           | 625.68             | 86                  | =0.05           |
| Xenon               | Xe                             | 131.36 | 289.74         | 5.838           | 1105.2             | 161.25              | 81.67           |
| Ethylene            | C <sub>2</sub> H <sub>4</sub>  | 28.05  | 282.65         | 5.075           | 216.7              | 104.0               | 0.10            |
| Ethane              | C <sub>2</sub> H <sub>6</sub>  | 30.07  | 305.88         | 5.010           | 217.59             | 89.9                | =0.5            |
| Carbon Dioxide      | CO <sub>2</sub>                | 44.01  | 304.19         | 7.383           | 464.00             | 216.54              | 517.3           |
| Sulfur Hexafluoride | SF <sub>6</sub>                | 146.05 | 318.69         | 3.759           | 736.0              | 222.35              | 223.9           |
| Propane             | C <sub>3</sub> H <sub>8</sub>  | 44.09  | 369.82         | 4.236           | 197.38             | 85.4                | -               |
| Ammonia             | NH <sub>3</sub>                | 17.03  | 406.80         | 11.627          | 237.64             | 195.4               | 6.077           |
| Isobutane           | C <sub>4</sub> H <sub>10</sub> | 58.12  | 409.07         | 3.685           | 194.51             | <sup>b</sup> 113.55 | na              |
| Water               | H <sub>2</sub> O               | 18.02  | 647.286        | 22.089          | 317.0              | 273.16              | 0.6113          |

<sup>a</sup>Lower Lambda Point of Helium    <sup>b</sup>Freezing Point

The structure of a thermodynamic state space of the kind shown in Fig.1 is constrained by the Gibbs phase rule which places a requirement on the number of independent variables required to determine state properties. The number of variables must equal the number of components present less the number of phases present plus two. In this single-component example, single-phases require two variables for their description and a 2-D diagram. Where two phases are in contact, only a single variable is required, and so such regions appear on the diagram as lines described by a relation between pressure and temperature. Likewise, all three phases can be in equilibrium only at one point in the *P-T* plane.

If the substance represented in the diagram were a mixture rather than a single component, an additional variable would be required to specify the state, and the saturation line together with the critical point would vary with composition. Also if the system were to include, e.g. elastic or electromagnetic forces and energy, quantities such as stress and strain or fields and permeabilities, would be required. It is essential in considering any thermodynamic problem to have a clear idea of the state space of the system involved, and to choose an appropriate set of variables to describe it.

**Thermophysical properties** Thermodynamic properties are the properties of equilibrium states

and include molecular weight, pressure-volume-temperature information, entropy, internal energy, enthalpy and free energy, and their derivatives. Additional properties are needed to describe certain systems. Derivative properties include the heat capacities, isothermal compressibility, and thermal expansivity. From these can be derived many additional measurable quantities such as latent heats, sound velocity, Joule-Thompson coefficient, heat capacity ratio etc.

Extension of thermodynamic formalism from the treatment of equilibrium states to processes occurring in finite time involves the definition of local relationships between thermodynamic quantities that are taken to be the same as the relationship in equilibrium. Thus we can speak of temperature or concentration gradients and energy or mass fluxes described by local thermodynamic variables and related by conductivities. The important transport properties in simple systems are thermal conductivity, viscosity, and diffusion coefficients. These are often tabulated with the thermodynamic properties and together are termed thermophysical properties.

There is a vast primary literature of thermophysical properties, and if something very particular is required, the best alternative is to use the bibliographical services of a specialized library [18]. More easily used are the property compilations found in standard reference books [1, 2]

Table 2 Selected Thermophysical Properties of Fluids [4,8,13,19]

|                     | State | Press-ure | Tempe-ature | Density | Enthalpy | Entropy | Cp                 | Thermal Cond. | Viscosity |
|---------------------|-------|-----------|-------------|---------|----------|---------|--------------------|---------------|-----------|
|                     |       | kPa       | K           | kg/cu-m | kJ/kg    | kJ/kg-K | kJ/kg-K            | mW/m-K        | microPa-s |
| Helium              | liq   | 101       | 4.222       | 124.9   | 10.02    | 3.575   | 13.58              | 18.7          | 3.263     |
|                     | gas   | 4.222     |             | 16.84   | 30.74    | 8.473   | 9.144              | 9.04          | 1.242     |
|                     | gas   | 300       |             | 0.163   | 1574     | 31.58   | 5.193              | 156.          | 19.92     |
| Hydrogen (para)     | liq   | 101       | 20.28       | 70.79   | -227.68  | 8.091   | 9.666              | 104.1         | 13.72     |
|                     | gas   | 20.28     |             | 1.339   | 217.75   | 30.058  | 12.241             | 14.68         | 1.109     |
|                     | gas   | 300       |             | .0818   | 4227.9   | 64.979  | 14.846             | 191.6         | 8.952     |
| Neon                | liq   | 101       | 27.09       | 1205    | -59.566  | 1.569   | 1.938              | 125.4         | 124.1     |
|                     | gas   | 27.09     |             | 9.592   | 26.145   | 4.733   | 1.179              | 7.52          | 4.735     |
|                     | gas   | 300       |             | 0.8193  | 309.031  | 7.253   | 1.030              | 49.36         | 31.59     |
| Nitrogen            | liq   | 101       | 77.31       | 806.8   | -121.872 | 2.832   | 2.042              | 133.0         | 150.1     |
|                     | gas   | 77.31     |             | 4.624   | 76.992   | 5.403   | 1.343              | 7.565         | 5.278     |
|                     | gas   | 300       |             | 1.138   | 311.405  | 6.842   | 1.041              | 25.78         | 17.95     |
| Argon               | liq   | 101       | 87.28       | 1394    | -117.599 | 1.372   | 1.120              | 128.3         | 260.3     |
|                     | gas   | 87.28     |             | 5.777   | 43.740   | 3.220   | 0.570              | 5.857         | 7.251     |
|                     | gas   | 300       |             | 1.624   | 156.084  | 3.876   | 0.522              | 17.91         | 22.90     |
| Oxygen              | liq   | 101       | 90.19       | 1141    | -133.107 | 2.942   | 1.699              | 151.8         | 194.8     |
|                     | gas   | 90.19     |             | 4.467   | 79.943   | 5.304   | 0.971              | 8.459         | 6.697     |
|                     | gas   | 300       |             | 1.300   | 272.970  | 6.413   | 0.920              | 26.29         | 20.63     |
| Methane             | liq   | 101       | 111.63      | 422.6   | -287.761 | 4.929   | 3.463              | 188.6         | 117.1     |
|                     | gas   | 111.63    |             | 1.817   | 222.572  | 9.501   | 2.207              | 11.92         | 4.462     |
|                     | gas   | 300       |             | 0.6528  | 625.968  | 11.616  | 2.236              | 34.13         | 11.24     |
| Krypton             | liq   | 101       | 119.7       | 2414    | -79.194  | 0.8236  | 0.466              | 88.86         | 399.8     |
|                     | gas   | 119.7     |             | 8.807   | 28.738   | 1.725   | 0.272              | 4.017         | 11.19     |
|                     | gas   | 300       |             | 3.411   | 74.409   | 1.959   | 0.249              | 9.473         | 24.92     |
| R-14                | liq   | 101       | 145.2       | 1633    | 18.08    | 0.1335  | <sup>a</sup> 0.942 | 123.5         | 490       |
|                     | gas   | 145.2     |             | 7.74    | 152.29   | 1.058   | 0.507              | =4.4          | 9.69      |
|                     | gas   | 300       |             | 3.59    | 245.27   | 1.485   | 0.705              | 15.82         | 17.1      |
| Xenon               | liq   | 101       | 165.04      | 2942    | -70.992  | 0.6076  | 0.363              | 70.72         | 490.5     |
|                     | gas   | 165.04    |             | 10.01   | 25.424   | 1.192   | 0.173              | 3.194         | 13.65     |
|                     | gas   | 300       |             | 5.362   | 47.496   | 1.290   | 0.160              | 5.567         | 23.12     |
| Ethylene            | liq   | 101       | 169.2       | 568.2   | -284.456 | 4.199   | 2.393              | 155.2         | 170       |
|                     | gas   | 169.2     |             | 2.089   | 197.855  | 7.047   | 1.297              | 3.6           | 5.82      |
|                     | gas   | 300       |             | 1.146   | 377.875  | 7.829   | 1.544              | 21.0          | 10.18     |
| Ethane              | liq   | 101       | 184.5       | 544.1   | -274.013 | 4.221   | 2.430              | 142.1         | 170       |
|                     | gas   | 184.5     |             | 2.056   | 214.471  | 6.869   | 1.437              | =3.3          | 5.95      |
|                     | gas   | 300       |             | 1.231   | 396.195  | 7.626   | 1.765              | 21.3          | 9.31      |
| Carbon Dioxide      | sol   | 101       | 194         | 1562    | 151.706  | 1.568   | <sup>a</sup> 1.244 | -             | na        |
|                     | gas   | 194       |             | 2.826   | 722.583  | 4.504   | 0.95               | 8.5           | 9.89      |
|                     | gas   | 293.15    |             | 1.839   | 803.579  | 4.835   | 0.843              | 16.9          | 14.68     |
| Sulfur Hexafluoride | liq   | 2103      | 293.15      | 1391    | 61.159   | 0.2293  | <sup>a</sup> 1.198 | -             | 291       |
|                     | gas   | 2103      | 293.15      | 191.1   | 127.932  | 0.4570  | 1.065              | -             | -         |
|                     | gas   | 101       | 293.15      | 6.145   | 147.854  | 0.6823  | 0.665              | 13.6          | 15.3      |
| Propane             | liq   | 101       | 231.1       | 581.2   | -199.314 | 3.881   | 2.247              | 129.3         | 220       |
|                     | gas   | 231.1     |             | 2.419   | 226.081  | 5.722   | 1.459              | 11.37         | 6.3       |
|                     | gas   | 300       |             | 1.820   | 334.495  | 6.131   | 1.704              | 17.97         | 8.0       |
| Ammonia             | liq   | 101       | 239.82      | 681.9   | -1497.97 | 5.123   | 4.472              | 620           | 255       |
|                     | gas   | 239.82    |             | 0.8901  | -129.11  | 10.831  | 2.348              | 18.81         | 8.12      |
|                     | gas   | 300       |             | 0.6992  | 3.93     | 11.326  | 2.166              | 24.56         | 9.9       |
| Isobutane           | liq   | 101       | 261.53      | 593.7   | -121.52  | 3.458   | 2.245              | 112.1         | 228.3     |
|                     | gas   | 261.53    |             | 2.827   | 245.15   | 4.860   | 1.561              | 12.57         | 6.642     |
|                     | gas   | 300       |             | 2.423   | 307.90   | 5.084   | 1.707              | 16.15         | 7.561     |
| Water               | liq   | 101       | 373.16      | 958.4   | -3174.85 | 10.646  | 4.215              | 690           | 283.8     |
|                     | gas   | 373.16    |             | 0.5975  | -918.29  | 16.694  | 2.073              | 23.9          | 12.6      |
|                     | liq   | 300       |             | 996.6   | -3481.36 | 9.732   | 4.180              | 620           | 857.4     |

<sup>a</sup>Heat Capacity at constant saturation

and handbooks [6, 7, 8]. These are good sources for tabulated fluid properties. There are, in addition, correlations and highly developed methods for the estimation of properties [3].

For intensive uses where tabulated material is not convenient, a computer-based property formulation is needed. A minimum form for this consists of a mechanical equation of state,  $P = P(\rho, T)$  correlating  $P$ - $\rho$ - $T$  data, a thermal equation of state correlating the ideal gas heat capacity as a function of temperature, a correlation of vapor pressure as a function of saturation temperature, and saturated liquid density correlated as a function of saturation temperature. From these four relations all of the thermodynamic properties can be calculated over the range of the correlations employed [4, 14]. In many cases  $P$ - $\rho$ - $T$  data for the saturated liquid only is available, and the mechanical equation of state may not be extended to include the liquid. This is the situation for correlations of the properties of many refrigerants and industrial gases [4, 13].

With a  $P$ - $\rho$ - $T$  correlation including both saturated and unsaturated liquid and gas, both the vapor pressure and the saturated densities may be found from the mechanical equation of state, and separate vapor pressure and liquid density correlations become redundant. These are often retained, however, for convenience of use. Care must be used in using tabulated data from multiple sources, e.g., a vapor pressure correlation and a table of latent heat, that may not be consistent. In fact, the laws of thermodynamics place requirements on the correlation of thermodynamic properties [16], and there is value added in the formulation of a property correlation that utilizes the full range of measured quantities.

Consistent correlations of thermodynamic and transport properties have been made for the air gases, the noble gases, most of the important refrigerants, and many hydrocarbons. These are available in computerized form together with the mathematical apparatus for calculating the full range of properties using a variety of independent variables [11, 12, 15, 17, 18, 19]. The use of such "thermodynamics packages" is very much recommended in all but the simplest applications.

**Chlorofluorocarbon gases** According to international agreement, the Montreal Protocol (1987), chlorofluorocarbon refrigerants, solvents, and fire suppression compounds are listed for their capacity to damage atmospheric ozone. Restrictions

on their manufacture are instituted, and plans for their replacement developed and implemented [20]. Any contemplated use of these substances should take into account their potential for environmental damage and possible restrictions on their availability. For a treatment of some replacements for these fluids see Sec.7.2.9.

## References

- [1] CRC Handbook of Chemistry and Physics, CRC Press, 75th ed. (1994)
- [2] R.H. Perry et al, Perry's Chemical Engineer's Handbook, McGraw-Hill (1984)
- [3] R.C. Reid et al, The Properties of Liquids and Gases, McGraw-Hill, 4th ed. (1987)
- [4] W.C. Reynolds, Thermodynamic Properties in SI, Dept. Mechanical Engineering, Stanford U. (1979)
- [5] ASHRAE Thermodynamic Properties of Refrigerants, American Society of Heating, Refrigerating, and Air-Conditioning Engineers, 345 E. 47th Street, New York, NY 10017 (1980)
- [6] Compressed Gas Association Handbook of Compressed Gases, Van Nostrand Reinhold, 3rd ed. (1990)
- [7] W. Braker, A.L. Mossman, Matheson Gas Data Book, Matheson Division of Searle Medical Products USA, Inc. Lyndhurst, NJ, 6th ed. (1980)
- [8] The Matheson Unabridged Gas Data Book, Matheson Division of Searle Medical Products USA, Inc. Lyndhurst, NJ
- [9] N.V. Vargaftik, Tables on the Thermophysical Properties of Liquids and Gases, Hemisphere Pub., 2nd ed. (1975)
- [10] National Reference Data Service of the USSR: A series of Property Tables, Hemisphere Pub.
- [11] C. Borgnakke, R.E. Sonntag, Thermodynamic and Transport Properties, Wiley (1997)
- [12] R.T. Jacobsen et al, Thermodynamic Properties of Cryogenic Fluids, Plenum (1997)
- [13] W.H. Mears et al, J. Phys. Chem. V73 (1969) 2254
- [14] J.G. Hust, A.L. Gosman, Advances in Cryogenic Engineering, V9, Plenum (1964) p.227
- [15] W.M. Haynes, D. G. Friend, Advances in Cryogenic Engineering, V39b, Plenum (1994) p.1865
- [16] R.T. Jacobsen et al, Advances in Cryogenic Engineering, V39b, Plenum (1994) p.1875
- [17] E.W. Lemmon et al, Advances in Cryogenic Engineering, V39b, Plenum (1994) p.1891
- [18] United States National Institute for Standards and Technology, The catalog of Standard Reference Data Products is available on the web

- at <<http://www.nist.gov/srd/>>, NIST Virtual Library at <<http://nvl.nist.gov/>>
- [19] HEPAK and GASPAK Computer Property Packages, Published by Cryodata, Inc., available from HORIZON Technologies, <http://www.htess.com/software.htm>
- [20] Montreal Protocol at <<http://www.unep.ch/ozone/Publications/MP-Handbook/>>; US Navy CFC and Halon Clearinghouse at <http://homer.ornl.gov/nuclearsafety/env/ods/reports.html>; US Environmental Protection Agency, Clean Air Act, Significant New Alternatives Policy (SNAP) <http://www.epa.gov/ozone/strathome.html>  
For general references:
- [21] E.B. Smith, Basic Chemical Thermodynamics, Oxford U., 4th ed. (1990)
- [22] R.A. Albery, R.J. Silbey, Physical Chemistry, Wiley (1992)
- [23] A. Bejan, Advanced Engineering Thermodynamics, Wiley (1988)

## 5.4 CREEP AND STRESS RELAXATION IN ACCELERATOR COMPONENTS

*F. Markley, FNAL*

The most used expression for fitting creep data is

$$\epsilon(t) = \epsilon_0 + \epsilon_1 t^n \quad (1)$$

where  $\epsilon(t)$  is the total deformation and  $\epsilon_0$  is the time independent deformation. Dividing by the constant stress  $\sigma_0$  gives the creep compliance

$$J(t) = J_0 + J_1 t^n \quad (2)$$

Creep is more commonly measured than stress relaxation because of the ease of maintaining a constant load. However stress relaxation is a closer approximation to common accelerator situations such as that of preloaded (collared) superconducting coils. The time independent strain and stress are inversely related, but the total stress and strain are related by the convolution integral

$$\int G(\tau) J(t - \tau) d\tau = t \quad (3)$$

which can be solved, with a limiting approximation, to give a stress relaxation modulus of the form

$$G(t) = G_0 - G_1 t^n \quad (4)$$

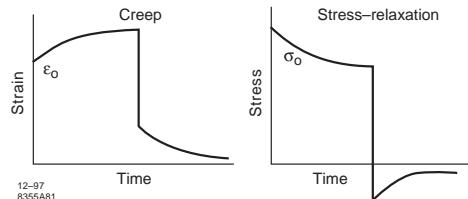


Figure 1: Generalized creep and stress relaxation curves.

which multiplied by the constant deformation  $\epsilon_0$  gives the time dependent stress

$$\sigma(t) = \sigma_0 - \sigma_1 t^n \quad (5)$$

Assuming time-temperature superposition, the relaxation modulus can be written as

$$G(t) = G_0 - G_1 (kt)^n \quad (6)$$

where  $k$  is the time-temperature shift factor. The time to relax a certain amount at a low temperature  $T_0$  is  $k$  times as long as the time needed at some higher temperature  $T_1$ . If the relaxation process has a characteristic time constant that can be described by an Arrhenius equation the shift factor can be written as

$$k = \exp \left[ Q \left( \frac{1}{T_0} - \frac{1}{T_1} \right) \right] \quad (7)$$

The relaxing material must not have a second order transition in the pertinent temperature interval. At zero temperature the shift factor becomes infinite and no relaxation occurs. Thus the time independent relaxation modulus  $G_0$  can be approximated as the elastic modulus of the material measured at a very low temperature. If relaxation measurements taken at two different temperatures are plotted as  $\ln(G_0 - G)$  then  $k$  can be obtained from the shift along the  $\ln t$  axis required to make the high temperature data appear as an extension of the low temperature data.

Fig.1 (left) shows a generalized creep curve and Fig.1 (right) shows a generalized stress relaxation curve, including recovery after removal of the initial strain or stress. Fig.2 shows some actual stress relaxation data for a simulated collared superconducting coil.

In general, soft materials like plastics relax faster than hard materials like metals. Cross-linked plastics like the epoxy resins commonly used to assemble superconducting coils may relax to some constant value where the cross-links have taken up a portion of the stress. However,

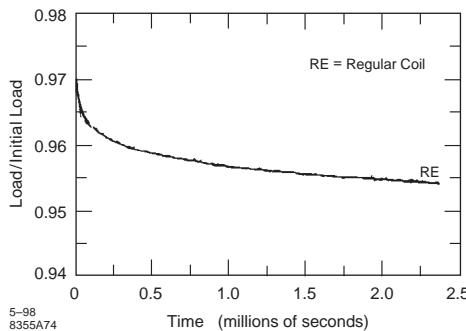


Figure 2: Example of stress relaxation.

thermoplastics like the polyimide often used as insulation on superconducting coils may relax indefinitely due to visco-elastic flow.

Superconducting coils are usually made of flattened cable, where the strands of one layer cross the strands of the adjacent layer at an angle, producing areas of “point” contact. The copper in these areas must deform until the stress is reduced to the yield stress because of increasing area. Under such a large stress, the copper may relax appreciably. Contrariwise, the film insulation between the coil layers is “spot” loaded and surrounded by film under much lower stress which can prevent its deformation, much like the retaining ring around the anvils in a high pressure apparatus. Thus the relaxation in the plastic is very much lower than expected.

As a measurement guide, it can be noted that Ref.[9] found  $n = 0.03$  for an epoxy fiberglass sample and Ref.[17] found  $Q = 5000$  for another epoxy fiberglass. Ref.[10] found  $Q = 11400$  and  $n = 0.08$  for a simulated superconducting coil.

## References

- [1] F. Garofalo, Fundamentals of Creep and Creep Rupture in Metals, Macmillan (1965)
- [2] F.J. Lockett, Nonlinear Viscoelastic Solids, Academi (1972)
- [3] I.L. Hopkins and R.W. Hamming, J. App. Phys. 28 (1957) 206
- [4] J.D. Ferry, Viscoelastic Properties of Polymers, Wiley (1980)
- [5] P.D. Ritchie, Physics of Plastics, D. Van Noststrand Co. (1965)
- [6] I.M. Ward, Mechanical Properties of Solid Polymers, Wiley (1971)
- [7] J. Schultz, Polymer Materials Science, Prentice-Hall (1974)
- [8] J.J. Alkonis, W.J. MacKnight, Introduction to Polymer Viscoelasticity, 2nd ed., Wiley-Interscience (1983)
- [9] F.W. Markley, FNAL Technical Support, Engineering Group, Internal Report
- [10] J.A. Carson, F.W. Markley, Mechanical Properties of Superconducting Coils (1984)
- [11] F.W. Markley, J.S. Kerby, Super Collider 2, Vol.2, Plenum (1990) p.753
- [12] F.W. Markley, O. Corpus, Super Collider 3, Vol.3, Plenum (1991) p.155
- [13] D. Rogers, F. Markley, Super Collider 4, Plenum (1992) p.351
- [14] J. McAdams, F. Markley, ibid. p.359
- [15] A. Lautenschlager et al, ibid. p.1031
- [16] R. Sims et al, Super Collider 5 (1993) p.839
- [17] E.M. Caulfield, PhD Thesis, U. of Ill. (1979)
- [18] DuPont Bulletin E-42727 on Kapton props. (1981)

## 5.5 ELECTRIC AND MAGNETIC FORCES

*M. Tigner, Cornell U.*

**Electric forces [1, 2]** 1. (a) Capacitance  $C$ , with constant  $q$ ;  $x$  is plate separation; motion reduces stored energy,

$$F_{\text{plate}} = \frac{q^2}{2} \frac{1}{C^2} \frac{dC}{dx} \quad (1)$$

(b) with constant  $V$ , motion increases stored energy,

$$F_{\text{plate}} = \frac{V^2}{2} \frac{dC}{dx} \quad (2)$$

2. (a) Internal pressure in dielectric, outward from dielectric,

$$P_{\text{out}} = \frac{\epsilon' - \epsilon_0}{2} E^2 \quad (3)$$

(b) Boundary pressure between two dielectrics 1 and 2 (towards medium with lower  $\epsilon$ ),

$$P_{\text{bound}} = \frac{1}{2} \left[ E_{1,\text{tang}}^2 (\epsilon'_1 - \epsilon'_2) + D_{1,\text{norm}}^2 \left( \frac{1}{\epsilon'_2} - \frac{1}{\epsilon'_1} \right) \right] \quad (4)$$

(c) Conductor-insulator boundary (normal towards insulator)

$$P_{\text{bound}} = \frac{D_{\text{norm}}^2}{2\epsilon'} \quad (5)$$

3. Torque on a dipole of dipole moment  $\vec{p}$ ,

$$\vec{\tau} = \vec{p} \times \vec{E} \quad (6)$$

**Magnetic forces [1, 2]**

## Current carrying elements

1. (a) On volume distribution of current in time constant field,

$$\vec{F} = \int_V \vec{J} \times \vec{B} dV \quad (7)$$

(b) On current  $I$  in thin wire ( $d\vec{s}$  in direction of  $I$ )

$$d\vec{F} = I d\vec{s} \times \vec{B} \quad (8)$$

(c) On parallel wires, separation  $d$ ,

$$F = \frac{\mu_0}{2\pi} \frac{i_1 i_2}{d} \quad (9)$$

(d) Torque on a magnetic dipole,

$$\vec{\tau} = \vec{m} \times \vec{B}, \quad \vec{m} = IA\hat{n} \quad (10)$$

where  $A$  is the loop area and  $\hat{n}$  is the unit vector normal to the loop.

2. Between two coils, mutual inductance  $M$ , in an arbitrary direction  $ds$ ,

$$F_s = \pm i_1 i_2 \frac{dM}{ds}, \quad \text{and} \quad \tau_s = \pm i_1 i_2 \frac{dM}{d\theta_s} \quad (11)$$

Mutual inductance is here defined as the incremental change in flux linkages in one coil due to an incremental change in current in the other, i.e.  $M \equiv nd\Phi/di$  where  $n$  is the number of turns in the first coil.

3. (a) Boundary pressure between materials of different permeability,

$$P = \frac{1}{2} \left[ H_{1,\tan}^2 (\mu_1 - \mu_2) + B_{1,\text{norm}}^2 \left( \frac{1}{\mu_2} - \frac{1}{\mu_1} \right) \right] \quad (12)$$

towards medium with lower permeability. (b) At air-iron boundary.

$$P = \frac{1}{2} \frac{B^2}{\mu_{\text{iron}}} \quad (13)$$

A contemporary accelerator application to the problem of focusing and bending magnets in the field of a detector solenoid can be found in [3].

**Eddy current forces** [4] In a changing dipole magnetic field (particularly e.g. in a quenching SC magnet), the equatorial eddy current pressure is

$$P = \sigma_c B \dot{B} b t \quad (14)$$

$$= 1.45 \times 10^{-4} [\text{psi}] B [\text{T}] \dot{B} \left[ \frac{\text{T}}{\text{s}} \right] b [\text{m}] t [\text{m}]$$

where  $b$  and  $t$  are the inner radius and the thickness of the round beam pipe,  $\sigma_c$  is its conductivity. The eddy current Joule heating power per linear length in a changing dipole field is  $\mathcal{P}/L = \pi \sigma_c b^3 t \dot{B}^2$ . In a quenching SC quadrupole, the

eddy current pressure is smaller by a factor  $\sim (Kb/B)^2$ , where  $K$  in T/m is the quad strength. If the pipe thickness is not uniform but has a variation  $t(1 + \epsilon \sin 2\theta)$  ( $|\epsilon| \ll 1$ ), the beam pipe experiences a net torque per linear length which is given by  $(\pi b^2 \epsilon/2)$  times the eddy current pressure.

**Permanent magnetic material** For material with permanent magnetic dipole moment per unit volume  $\vec{M}$ , immersed in an external magnetic field  $\vec{B}$ , and assuming that any induced magnetization is negligible, the potential energy is [5]

$$W = - \int \vec{M} \cdot \vec{B} dv \quad (15)$$

Accordingly the forces acting on a piece of this material will be

$$\vec{F} = \nabla W \quad (16)$$

An important contemporary application is found in the engineering of PM multipoles and insertion devices where the large forces among the components require special precautions [6] and tooling for both safety and stability of the mechanical structure. The forces are typically 100's of pounds. (See also Sec.7.2.8.)

**References**

- [1] S.S. Attwood, Electric and Magnetic Fields, Wiley (1949)
- [2] J.A. Stratton, Electromagnetic Theory, McGraw-Hill (1941)
- [3] J.J. Welch, G. Dugan, Forces on Interaction Region Quadrupoles and Dipoles Due to a Detector Solenoid Magnet, CBN 97-22 Cornell U. (1997), <http://www.lns.cornell.edu/public/CBN/1997/CBN97-22/>
- [4] A. Chao, SSC-N-434 (1987)
- [5] D.E. Gray, ed., AIP Handbook, 5-36, AIP (1972)
- [6] W. Lou et al, PAC 97 and CBN97-4, Cornell U. (1997), <http://www.lns.cornell.edu/public/CBN/1997/CBN97-4/>

**5.6 DEFLECTIONS AND BUCKLING**

*M. Tigner, Cornell U.*

**Deflections**

Fig.1 shows deflections in common situations.

|   | Slope at free end.  | Deflection at any section in terms of $x$ : $\delta$ is positive downward.   | Maximum deflection.  |
|---|---|--|--|
| 1. Cantilever Beam – Concentrated load $P$ at the free end.                 |   |  |  |
|   | $\theta = \frac{P\ell^2}{2EI}$  | $\delta = \frac{Px^2}{6EI} (3\ell - x)$  | $\delta_{max} = \frac{P\ell^3}{3EI}$   |
| 2. Cantilever Beam – Concentrated load $P$ at any point.                    |   |  |  |
|   | $\theta = \frac{Pa^2}{2EI}$   | $\delta = \frac{Px^2}{6EI} (3a - x) \text{ for } 0 < x < a$<br>$\delta = \frac{Pa^2}{6EI} (3x - a) \text{ for } a < x < \ell$  | $\delta_{max} = \frac{Pa^2}{6EI} (3\ell - a)$  |
| 3. Cantilever Beam – Uniformly distributed load of $w$ .                    |   |  |  |
|   | $\theta = \frac{w\ell^3}{6EI}$  | $\delta = \frac{wx^2}{24EI} (x^2 + 6\ell^2 - 4\ell x)$   | $\delta_{max} = \frac{w\ell^4}{8EI}$   |
| 4. Cantilever Beam – Uniformly varying load; maximum intensity $w$ .        |   |  |  |
|   | $\theta = \frac{w\ell^3}{24EI}$   | $\delta = \frac{wx^2}{120EI} (10\ell^3 - 10\ell^2 x + 5\ell x^2 - x^3)$  | $\delta_{max} = \frac{w\ell^4}{30EI}$  |
| 5. Cantilever Beam – Couple $M$ applied at the free end.                    |   |  |  |
|   | $\theta = \frac{M\ell}{EI}$   | $\delta = \frac{Mx^2}{2EI}$  | $\delta_{max} = \frac{M\ell^2}{2EI}$   |
| 6. Beam freely supported at the ends – Concentrated load $P$ at the center. |   |  |  |
|   | $\theta_1 = \theta_2 = \frac{P\ell^2}{16EI}$  | $\delta = \frac{Px}{12EI} \left( \frac{3\ell^2}{4} - x^2 \right) \text{ for } 0 < x < \frac{\ell}{2}$  | $\delta_{max} = \frac{P\ell^3}{48EI}$  |
|   | Slope at ends.  | Deflection at any section in terms of $x$ : $\delta$ is positive downward.   | Maximum and center deflections.  |
| 7. Beam freely supported at the ends – Concentrated load at any point.      |   |  |  |
|   | $R_1 = \frac{Pb}{q}$<br>$R_2 = \frac{Pa}{q}$<br>$\theta_1 = \frac{Pb(\ell^2 - b^2)}{6EI}$<br>$\theta_2 = \frac{Pab(\ell - b)}{6EI}$ | $\text{Left End: } \theta_1 = \frac{Pb(\ell^2 - b^2)}{6EI}$<br>$\text{Right End: } \theta_2 = \frac{Pab(\ell - b)}{6EI}$<br>$\text{To the left of load P: } \delta = \frac{Pbx}{6EI} (\ell^2 - x^2 - b^2)$<br>$\text{To the right of load P: } \delta = \frac{Pbx}{6EI} b (x-a)^3 + (\ell^2 - b^2)x - x^3$ | $\delta_{max} = \frac{Pb(\ell^2 - b^2)^{3/2}}{9\sqrt{3}EI}$<br>at $x = \sqrt{\frac{b^2 - b^2}{3}}$<br>At center, if $a=b$<br>$\delta = \frac{Pb}{48EI} (3\ell^2 - 4b^2)$ |
| 8. Beam freely supported at the ends – Uniformly distributed load of $w$ .  |   |  |  |
|   | $R_1 = \frac{w\ell}{2}$<br>$R_2 = \frac{w\ell}{2}$<br>$\theta_1 = \theta_2 = \frac{w\ell^3}{24EI}$                                  | $\delta = \frac{wx}{24EI} (x^3 - 2\ell x^2 + x^3)$   | $\delta_{max} = \frac{5w\ell^4}{384EI}$  |
| 9. Beam freely supported at the ends – Couple $M$ at the right end.         |   |  |  |
|   | $R_1 = \frac{M}{\ell}$<br>$R_2 = \frac{M}{\ell}$<br>$\theta_1 = \frac{M\ell}{6EI}$<br>$\theta_2 = \frac{M\ell}{3EI}$                | $\delta = \frac{M\ell x}{6EI} \left( 1 - \frac{x^2}{\ell^2} \right)$   | $\delta_{max} = \frac{M\ell^2}{9\sqrt{3}EI}$<br>at $x = \frac{\ell}{\sqrt{3}}$<br>At center<br>$\delta = \frac{M\ell^2}{16EI}$   |
| 10. Beam freely supported at the ends – Couple $M$ at the left end.         |   |  |  |
|   | $R_1 = \frac{M}{\ell}$<br>$R_2 = \frac{M}{\ell}$<br>$\theta_1 = \frac{M\ell}{6EI}$<br>$\theta_2 = \frac{M\ell}{3EI}$                | $\delta = \frac{Mx}{6EI} (\ell - x)(2\ell - x)$  | $\delta_{max} = \frac{M\ell^2}{9\sqrt{3}EI}$<br>at $x = (1 - \frac{1}{\sqrt{3}})\ell$<br>At center<br>$\delta = \frac{M\ell^2}{16EI}$                                    |

Figure 1: Deflections for some common beam configurations.  $E$  = Young's modulus;  $I$  = "moment of inertia". The moment of inertia of a plane area about the  $x$ -axis is  $I_x = \int_A y^2 dA$ . Above  $x$ -axis = neutral axis of beam.

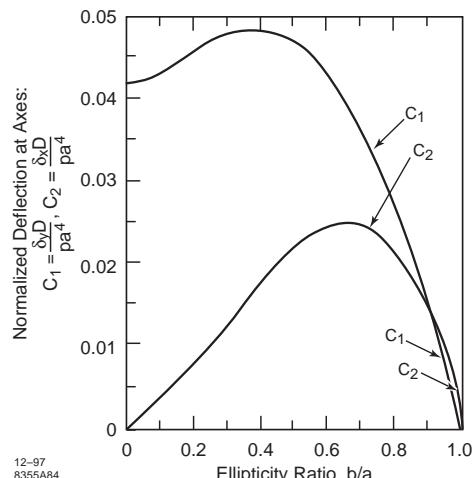


Figure 2: Deflections of elliptical beam pipe.

Waveguide deflection For internal or external pressure the deflection of the broad wall of 2:1 rectangular waveguide is given by [2], in SI units

$$\delta = 0.065 \frac{PW^4}{Et^3} \quad (1)$$

with  $P$  the gauge pressure,  $W$  the broad wall width,  $t$  the wall thickness,  $E$  the Young's modulus of the guide material.

Elliptical vacuum chamber deflection For an elliptical vacuum chamber of semi-minor axis  $b$  and semi-major axis  $a$ , uniform thickness  $t$ , subjected to uniform gauge pressure  $P$ , the deflections [3] at the extrema of the minor and major axes are

$$\delta_y = C_1 Pa^4/D; \quad \delta_x = C_2 Pa^4/D \quad (2)$$

where  $D = Et^3/12(1 - \mu^2)$

and the  $C$ 's are given in Fig.2.  $\mu$  is Poisson's ratio (0.25-0.35 metals, 0.1 for some concretes and up to 0.5 for rubber).

**Buckling formulae [4]** (a) The critical load for column, length  $L$ , buckling  $W_{cr}$  is given by [4]

$$W_{cr} = K \frac{\pi^2 EI}{L^2} \quad (3)$$

$K$  depends upon end conditions:

free-free [ $K = 1$ ], free-guided [0.25]; clamped-free [0.25]; free-pinned [1]; pinned-pinned [1]; clamped-pinned [2.05]; clamped-clamped [4]; clamped-guided [1]; guided-pinned [0.25]; guided-guided [1].

(b) For a thin walled cylinder thickness  $t$ , external pressure  $P$ , length  $L$ , and radius  $R$ , the critical pressure for collapse is

$$P_{cr} = \sigma_{cr} \frac{t}{R}, \quad \text{where} \quad (4)$$

$$\sigma_{cr} = K_c \frac{\pi^2 E}{12(1 - \mu^2)} \left( \frac{t}{R} \right)^2, \quad K_c = 10^K$$

$$K = 0.6337 - 0.1455 \log_{10} Z + 0.1977 \log_{10}^2 Z - 0.019151 \log_{10}^3 Z$$

$$Z = \left( \frac{L^2}{Rt} \right) \sqrt{(1 - \mu^2)}$$

One may take into account, approximately, the effects of non-uniformity of  $t$  and  $R$  by multiplying the critical pressure by the factor  $(\hat{R}t/\check{R}t)^3$  where  $\hat{\cdot}$ ,  $\check{\cdot}$  refer to max and min values respectively.

(c) The critical buckling torque for a cylindrical shell, variables as in (b), is

$$T_{cr} = 2\pi R^2 K_\tau \frac{t^2}{RZ^{1/4}}, \quad \text{where} \quad (5)$$

$$K_\tau = 0.4233 + 79.978 \eta^{-1} - 12.76 \times 10^3 \eta^{-2} + 7.56 \times 10^5 \eta^{-3}, \quad \eta = R/t$$

**Properties of beryllium** Beam tubes of very low  $Z$  are frequently made from Beryllium. Its mechanical properties relevant to beam tube service are:  $E = 290$  GPa, Yield varies 200-400 MPa.  $\mu = 0.24-0.3$ , therm exp.  $11.6 \times 10^{-6}/^\circ\text{C}$ , sp. grav. 1.85, thermal conduc. 146 W/m-K, spec. heat 1886 J/kg, elec. resistivity at  $22^\circ\text{C}$ ,  $3.25 \times 10^{-8}$ , temp. coeff of resistivity,  $0.025 \times 10^{-8}\Omega \cdot \text{m}/^\circ\text{C}$ .

## References

- [1] S. Timoshenko, G. MacCullough, Elements of the Strength of Materials, D. van Nostrand (1935)
- [2] T.S. Saad, ed., Microwave Engineers Handbook, Artech House (1971)
- [3] R. Avery, G. Tidrick, PAC 69, p.952
- [4] W.D. Pilkey, Formulas for Stress, Strain and Structural Matrices, Wiley (1994)

## 5.7 PRACTICAL HEAT TRANSFER AND FLUID FLOW

M. McAshan, FNAL

M. Tigner, Cornell U.

The principal modes of heat transfer are conduction, convection and radiation. Typical heat transfer situations by type, surface temperature and specific dissipation are displayed in Tab.1 [1].

Table 1

| Type         | T[°C]    | q[W/cm <sup>2</sup> ] |
|--------------|----------|-----------------------|
| radiation    | 400-1000 | 4-10                  |
| water        | 30-150   | 30-110                |
| forced air   | 150-200  | 0.5-1                 |
| evaporative* | 100-120  | 80-125                |
| conduction   | 100-250  | 5-30                  |

\* static pool, water

**Heat transfer by conduction** Heat conduction is the process of diffusion of molecular motion in response to a gradient in temperature, and it is described by the two equations of Fourier:

$$\vec{q} = k\nabla T \quad \text{and} \quad \nabla \cdot \vec{q} = \rho c \frac{\partial T}{\partial t} - \rho_s \quad (1)$$

The first of these relates the energy flux density to the gradient of the temperature field.  $k$  is the thermal conductivity which may be a function of position and/or of temperature [8]. The second equation expresses the conservation of energy, relating the divergence of the energy flux to the rate of change of the local energy density less any source density,  $\rho_s$ . The quantity  $\rho c$  is heat capacity per unit volume. Note that fluids can be heated at constant pressure or at constant volume, and while most problems involving solids are at constant pressure, significant strain energy can be present. Materials properties are available tabulated [3, 4, 6, 11, 13, 17] or in the form of computer correlations [18, 19]. See Sec.5.1.

Many methods [6, 8, 16] have been developed for the solution of these equations. For steady-state problems with constant  $k$ , the temperature field satisfies Poisson's or Laplace's equation. Thus the same methods that apply in electro- and magnetostatics and hydrodynamics apply here [8]. Many computer-aided engineering codes such as ANSYS® have solvers for steady-state thermal problems in full 3-D.

The assumption of constant  $k$  is not restrictive because the form of Fourier's equations remains the same if the temperature variable is replaced with  $\lambda(T)$ , the thermal conductivity integral [9]. Thus steady-state solutions for  $\lambda(T)$  can be taken directly from the same solutions with constant  $k$ . For the 1-D case the conductivity integral is particularly convenient. The heat flux is constant and  $\lambda$  is linear in the coordinate. Thus  $q = A[\lambda(T_2) - \lambda(T_1)]/L$  where  $A$  is the area of the conductor,  $L$  its length and the  $T$ 's the temperature at the two ends.

Transient solutions of Fourier's equations depend on the group  $\kappa = k/\rho c$  which is referred

to as the thermal diffusivity and has units [ $L^2/t$ ]. In transient thermal analysis it often happens that the characteristic times associated with the parts of the system will be much shorter than that associated with the interaction of parts. This leads to very useful approximation by a lumped-circuit model of thermal capacities linked by thermal resistances. Codes based on this approach are used for the thermal analysis of spacecraft and cryogenic systems. It can be useful to turn this process around and from measurements of thermal time-constants in a system infer an appropriate lumped model and its parameters.

**Heat transfer by convection** Convective heat transfer includes all of the processes of heat transfer to, from, and by fluid flow. For practical purposes the description of steady fluid flow is reduced to 1-D by averaging over velocity and temperature distributions. These distributions are referred to as the fully-developed profiles. The variables of the reduced system are mass-flow averaged velocity and thermophysical properties. In this regime, the flux of energy in the stream is

$$\text{Energy flux} = G \left[ H + \frac{v^2}{2} + g(z - z_0) \right] \quad (2)$$

and average coefficients for the transport of heat to and from the flow and for viscous friction are expressed in semi-empirical dimensionless relationships consistent with hydrodynamic similarity. Fluid conditions are described by any appropriate pair of variables such as  $T$  and  $P$ , or  $H$  and  $P$ . In many situations the velocity is low enough to omit kinetic energy from the description.

Steady duct flow is characterized by the Reynolds number alone. Three kinds of flow are observed: For Reynolds number  $< 2000$ , the flow is termed laminar and  $> 10,000$  turbulent; in the transition region the character is unstable with slowly-developing, large-scale eddies.

Pressure drops in duct flow are given by the expressions:

$$P_1 - P_2 = \frac{G^2}{2\rho} \left( 4f \frac{\Delta L}{d} + K \right) + \rho g \Delta z \quad \text{friction}$$

$$P_1 - P_2 = \frac{G^2}{2\rho} \left( 1 - \frac{A_2^2}{A_1^2} + K_c \right) \quad \text{contraction}$$

$$P_1 - P_2 = \frac{G^2}{2\rho} \left( \frac{A_1^2}{A_2^2} - 1 + K_e \right) \quad \text{expansion}$$

$$P_1 - P_2 = \frac{G^2}{2\rho_1} 2 \left( \frac{\rho_1}{\rho_2} - 1 \right) \quad \text{change of density}$$

## Sec.5.7: PRACTICAL HEAT TRANSFER AND FLUID FLOW

Table 2 Definitions  
(One dimensional fluid flow in a duct)

|                      |   |
|----------------------|---|
| $A, p$               | Area & perimeter of duct  |
| $d$                  | Hydraulic diameter = $4A/p$                                       |
| $G$                  | Mass flow per unit area, $\text{kg}/(\text{s} \cdot \text{m}^2)$  |
| $h$                  | Heat transfer coefficient, $\text{W}/(\text{m}^2 \cdot \text{K})$ |
| $K, K_{c,e}$         | Head loss (resistance) coefficient                                |
| $\rho, P, T, H$      | Density, Pressure, Temperature, Enthalpy                          |
| $k$                  | Thermal conductivity, $\text{W}/(\text{m} \cdot \text{K})$        |
| $C_p$                | Heat capacity, const. $P$ , $\text{J}/(\text{kg} \cdot \text{K})$ |
| $q', q''$            | Latent heats  |
| $\mu$                | Fluid dynamic viscosity, $\text{Pa} \cdot \text{s}$               |
| $\alpha$             | Fluid expansion coefficient - $(dp/dT)_p$ , $\text{K}^{-1}$       |
| $f$                  | Fanning friction factor   |
| $v$                  | Average fluid velocity - $G/d$ , $\text{m}/\text{s}$              |
| $z, g$               | Vertical height, acceleration of gravity                          |
| Dimensionless Groups |   |
| $Re_d$               | Reynolds number - $dg/\mu = dpv/\mu$                              |
| $Pr$                 | Prandtl number - $\mu C_p/k$                                      |
| $Nu_d$               | Nusselt number - $dh/k$   |
| $St$                 | Stanton number - $h/GC_p = Nu/(Re Pr)$                            |
| $Ra_x$               | Rayleigh number - $Pr \cdot x^3 \rho g \Delta T / \mu^2$          |

The frictional drop is proportional to a friction factor  $f$ . The Darcy friction coefficient, also called  $f$  and equal to  $4f$ , is sometimes used. Friction factor correlations are given in Tab.3. At high Reynolds numbers wall roughness becomes important in limiting the decrease of the friction factor, and a diagram, published first by Moody, gives friction factor as a function of Reynolds number and a relative roughness [3, 4, 5, 6, 15]. The content of the Moody diagram is represented by an implicit equation by Colebrook and White that is convenient for computer calculation [5, 15].

Frictional pressure drops in duct irregularities are represented by head-loss factors  $K$ . As an example, the loss due to a standard-radius 90 degree elbow is  $K = 0.75$ . Tabulations of  $K$  factors appear in [3, 4, 5] and are sometimes presented as equivalent  $\Delta L/d$ .

Pressure changes due to flow acceleration in changes of cross-section, changes of density as a function of temperature and changes in height ( $\rho g \Delta z$ ) are often important. Change in area also produces frictional loss, and  $K$  factors for these situations are tabulated [4, 5, 9].

The dimensionless heat transfer coefficient in fully-developed flow is the Nusselt number. This is correlated as a function of the Reynolds number and the Prandtl number which is a dimensionless fluid property. A selection of correlations

Table 3 Friction Factor and Heat Transfer Correlations in Fully Developed Flow [8,10]

|  |   |
|--|---|
| flow in a duct, hydraulic diameter $d$   |   |
| $f = 16/\text{Re}$   | laminar, $\text{Re}_d < 2000$                       |
| $f = 0.079/\text{Re}_d^{0.25}$   | $5000 < \text{Re}_d < 30,000$                       |
| $f = 0.046/\text{Re}_d^{0.2}$  | $30,000 < \text{Re}_d < 1,000,000$                  |
| $Nu_d = 4.36$  | laminar, round, $\text{Re} < 2000$                  |
| $Nu_d = 0.022\text{Pr}^{0.6}\text{Re}_d^{0.8}$   | gases, $0.5 < \text{Pr} < 1.0$                      |
| $Nu_d = 0.0155\text{Pr}^{0.5}\text{Re}_d^{0.83}$   | liquids, $1 < \text{Pr} < 20$                       |
| $Nu_d = 0.0118\text{Pr}^{0.3}\text{Re}_d^{0.9}$  | oils, $\text{Pr} > 20$                              |
| <i>laminar forced convection, velocity <math>v</math>, over plate length <math>x</math>: <math>x</math> average Nusselt Number</i> |   |
| $Nu_x = 0.664\text{Pr}^{1/3}\text{Re}_x^{1/2}$   | $0.6 < \text{Pr} < 50, \text{Re}_x < 5 \times 10^5$ |
| <i>natural convection over vertical plate, height <math>x</math>: <math>x</math> average Nusselt Number</i>                        |   |
| $Nu_x = 0.59\text{Ra}_x^{1/4}$   | laminar, $10^4 < \text{Ra}_x < 10^9$                |
| $Nu_x = 0.13\text{Ra}_x^{1/3}$   | turbulent, $10^9 < \text{Ra}_x < 10^{12}$           |

for ducted flow and for free and forced convection over a plate are given in Tab.3. Many additional correlations for non-fully developed flow situations in ducts and for more complicated external flow conditions are available [6, 7, 15].

**Cooling systems based on water** For heat removal by flowing water, the fundamental equation relating  $P$ , the power removed [W] to the volume flow rate,  $Q$  [g/s], and the temp. rise in the water  $\Delta T$  [K] is

$$Q = 0.24 P / \Delta T \quad [\text{g}/\text{s}] \quad (3)$$

Take as an example of this type of cooling a magnet conductor 10 m long with a dissipation of 1 kW per meter cooled by a flow of water in a 1 cm diameter channel. For a 5 K temperature rise,  $Q$  must be 7.6 gpm = 0.48 kg/s. Using the definitions and correlations above together with the properties for water at 20°C in Tab.4, we find:  $G = 6,112 \text{ kg}/(\text{s} \cdot \text{m}^2)$ ;  $Re = 60,515$ ;  $f = .0051$ ;  $Nu = 382$ ;  $h = 2.3 \times 10^4 \text{ W}/(\text{m}^2 \cdot \text{K})$ . Finally we find  $\Delta P = 380 \text{ kPa}$  and a temperature difference  $\Delta T = 1.38 \text{ K}$ .

**Natural convection cooling** Using the correlation in Tab.3 for heat transfer in the laminar region to a vertical panel of length  $X_L$  and a height  $X_H$  at a temperature  $\Delta T$  above the ambient air at 300 K, the power removed on one side by free convection is

$$P = 2.68 (T_R - T_A)^{5/4} X_H^{3/4} X_L \quad (4)$$

The units are meter, watt and K. Fig.1a&b shows the performance of a typical extruded aluminum heat sink used for transistor cooling. Plotted

Table 4 Thermophysical Properties of Water

| T<br>C | $\rho$<br>$\text{kg/m}^3$ | Cp<br>J/kg    | $\mu$<br>$\text{Pa-s}$ | k<br>$\text{W/m-K}$ | Pr   | $\alpha$<br>1/C  | $q'$<br>J/kg  |
|--------|---------------------------|---------------|------------------------|---------------------|------|------------------|---------------|
| mult   | $\times 10^3$             | $\times 10^3$ | $\times 10^{-4}$       |                     |      | $\times 10^{-4}$ | $\times 10^6$ |
| 0      | 1.00                      | 4.22          | 17.9                   | .569                | 13.3 | .691             | 2.50          |
| 20     | 1.00                      | 4.18          | 10.1                   | .603                | 7.00 | 2.08             | 2.45          |
| 40     | .992                      | 4.18          | 8.01                   | .632                | 5.30 | 3.86             | 2.41          |
| 60     | .983                      | 4.19          | 4.69                   | .654                | 3.00 | 5.24             | 2.36          |
| 80     | .972                      | 4.20          | 3.57                   | .670                | 2.24 | 6.42             | 2.31          |
| 100    | .958                      | 4.22          | 2.84                   | .680                | 1.76 | 7.50             | 2.26          |

Table 5 Thermophysical Properties of Air at Atmospheric Pressure

| T<br>K | $\rho$<br>$\text{kg/m}^3$ | Cp<br>J/kg | $\mu$<br>$\text{Pa-s}$ | k<br>$\text{W/m-K}$ | Pr   | $\alpha$<br>1/K  |
|--------|---------------------------|------------|------------------------|---------------------|------|------------------|
| mult   | $\times 10^3$             |            | $\times 10^{-6}$       | $\times 10^{-3}$    |      | $\times 10^{-3}$ |
| 100    | 3.56                      | 1.03       | 7.11                   | 9.34                | .786 | 10.8             |
| 200    | 1.75                      | 1.01       | 13.3                   | 18.1                | .737 | 5.06             |
| 300    | 1.16                      | 1.01       | 18.4                   | 26.3                | .707 | 3.34             |
| 400    | .871                      | 1.01       | 23.0                   | 33.8                | .690 | 2.50             |
| 500    | .696                      | 1.03       | 27.0                   | 40.7                | .684 | 2.00             |
| 600    | .584                      | 1.05       | 30.6                   | 46.9                | .685 | 1.67             |
| 700    | .498                      | 1.08       | 33.9                   | 52.4                | .695 | 1.43             |

against  $X_H$  are curves of  $\Delta T/P$  for values of  $P$ . Taking  $X_L \approx 1$  m, and re-arranging the above expression we get  $\Delta T/P = 0.7/X_H^{3/5}P^{1/5}$ . A little work with a calculator shows that this expression gives accurate values of  $\Delta T/P$  for  $X_H = 0.1$  m but underpredicts at the 0.3 m height by about 20%. This suggests that for the longer length the boundary layer thickness at the top of the heat sink approaches the spacing between the fins, reducing the heat exchange rate in this area.

**Forced air cooling** For the ducted flow of air the heat capacity together with the ideal gas law give an equation connecting power removed,  $P$  [W] with  $Q_{\text{air}}$  [ $\text{m}^3/\text{min}$ ], the volume flow of air, the temperature and pressure of the intake air,  $T_1$  [K] and  $p$  [bar], and the temp. of the exhaust air,  $T_2$  [K], is

$$P = 5968 p Q_{\text{Air}} \left( \frac{T_2}{T_1} - 1 \right) \quad (5)$$

Pressure drops and heat transfer coefficients can be found using the correlations in Tab.3. For the laminar forced flow of air over a plate, the correlation gives for 300 K ambient:

$$P = 3.91 v^{1/2} X_H^{1/2} X_L \Delta T \quad (6)$$

Weight: 4 kg per 1.5 m  
Thermal capacitance:  $3.56 \times 10^3$  J/K per 1.5 m  
Dimensions in mm

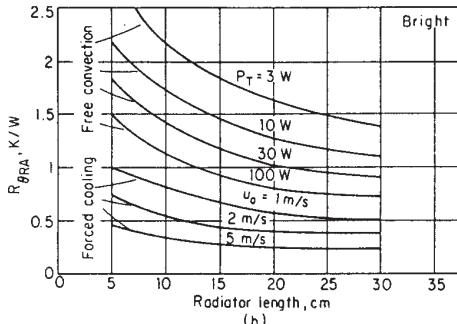
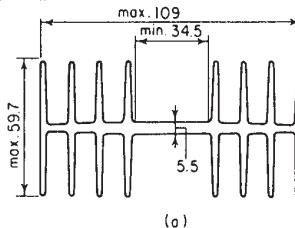


Figure 1: Al heat sink and performance.

Again with  $X_L \approx 1$  m, this expression describes the forced cooling curves in Fig.1b [2] with good accuracy.

**Heat exchangers** (See also Sec.5.9.1) A heat exchanger is two or more ducted flows arranged in thermal contact. The arrangement can be in counterflow, in cross-flow, or in parallel flow. Heat exchanger surfaces are described by friction factors and Stanton numbers given as functions of the Reynolds number, and by surface area per unit volume and hydraulic diameter [10]. Pressure drops include not only those associated with the heat exchange surfaces but also drops associated with distributors, piping, changes in flow area, flow acceleration and gravity. Pumping power is often an important factor in optimization.

Constant fluid properties are the basis of the design methods for heat exchangers [6, 7, 10]. These express the operation of the heat exchanger in terms of the inlet and outlet conditions and mass flow. As soon as variable fluid properties are involved, these methods can be misleading. It is necessary in this case to either integrate the equations with real fluid properties, or subdivide the

heat exchange path into small enough pieces that properties are constant over each piece.

Heat exchangers are specified by overall operating characteristics and mechanical requirements, and the detailed design is left to the manufacturer. If an extensive set of high performance heat exchangers is needed, it is wise to hire an experienced consulting engineer to prepare the requirements and specifications.

**Boiling heat transfer** Fig.2 illustrates the relation between a liquid-surface  $\Delta T$  and the pool boiling heat flux density [4, 6, 7, 14]. Three regimes of heat transfer can be seen: at low  $\Delta T$  heat transfer takes place through a natural convection process; and at somewhat higher  $\Delta T$  stable nucleate boiling occurs over several orders in heat flux. In this regime, the heat flux typically depends on  $\Delta T^{2.5}$ . At still higher  $\Delta T$ , nucleate boiling becomes unstable, the heat flux reaches a maximum, and under constant- $q$  conditions the  $\Delta T$  will jump to a much higher value in the film boiling regime. As indicated there is hysteresis in this behavior.

Under forced-flow conditions these regimes still occur, the vapor produced in the boiling being re-condensed or carried into the flow.

**Superfluid heat transport [12]** In superfluid helium, liquid helium below 2.2 K, a mode of heat transport takes place that has no counterpart in any other fluid. The heat flux  $q$  in the static fluid in a duct, length  $L$ , is characterized by small  $\Delta T/\Delta x$  below a peak heat flux  $q^*$ . Further, thermal flux  $q_s$  between the fluid and a surface has associated with it the Kapitza resistance (see also Sec.7.3.9), which produces a temperature difference between

the surface,  $T_s$  and the bath  $T_b$ .

$$\Delta T = f(T, p)q^m \Delta x : m = 3$$

$$f(T, p) > f(1.93, \text{sat}) = 7.4 \times 10^{-4} \text{cm}^5 \text{K/W}^3$$

$$q^* L^{1/m} = Z(T_b) : Z(1.8) = 7.5 \text{ W/cm}^2 \text{cm}^{1/3}$$

$$q_s = a(T_s^n - T_b^n) : n = 3.45; a = .0455 \text{ W/cm}^2 \text{K}$$

Here  $m$  can be taken as 3, and some typical or limiting values for  $f$  and  $Z$  are shown. The constant  $a$  and the exponent  $n$  are both material and temperature dependent. The values given are for polished copper at  $T_b = 1.8$  K.

**Heat transfer by radiation** Radiative power transfer, watts, from a surface, area  $A_1 [\text{m}^2]$  having an emissivity  $\epsilon_1$  at temp.  $T_1 [\text{K}]$ , radiating into a surface area  $A_2$  at temp.  $T_2$  is [6, 7, 16]

$$P_{12} = A_1 \sigma_{SB} (T_1^4 - T_2^4) / \left[ \frac{1}{\epsilon_1} + \frac{(1 - \epsilon_2)A_1}{\epsilon_2 A_2} \right] \quad (7)$$

where  $\sigma_{SB}$  is the Stefan-Boltzmann constant  $5.67 \times 10^{-8} \text{ W/m}^2 \text{K}$ . This form applies to the situation that the two surfaces form an enclosure, that all of the radiation leaving surface 1 falls on surface 2, and that all surfaces are diffuse gray bodies with emissivities  $\epsilon_1$  and  $\epsilon_2$ . This is a good assumption for most materials at room temperature and below, but can be untrue at higher temperatures. Some emissivities of materials commonly used in radiative cooling situations are shown in Tab.6 [1, 2].

Important special cases include a small body inside of a large ambient ( $A_1/A_2 \approx 0$ ) for which  $P_{12} = A_1 \epsilon_1 \sigma_{SB} (T_1^4 - T_2^4)$  and identical parallel planes ( $A_1/A_2 = 1, \epsilon_1 = \epsilon_2 = \epsilon$ ) where  $P_{12} = A_1 \epsilon \sigma_{SB} (T_1^4 - T_2^4)/2$ . Cases for cylindrical ( $A_1/A_2 = r_1/r_2$ ) and spherical geometries ( $A_1/A_2 = r_1^2/r_2^2$ ) can easily be derived.

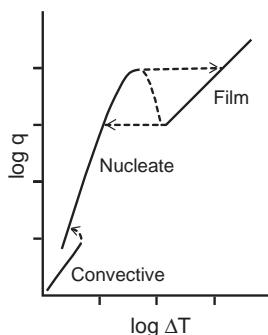


Figure 2: Typical pool boiling curve.

Table 6

| Mat'l          | T[K] | $\epsilon^*$ |
|----------------|------|--------------|
| Al(polished)   | 450  | 0.04         |
| Al(blk. anod.) | 315  | 0.5          |
| Graphite       | 1000 | 0.9          |
| Cu(bright)     | 300  | 0.07         |
| Cu(blk. oxide) | 300  | 0.8          |
| Mo             | 1300 | 0.13         |
| Ni             | 600  | 0.09         |
| Ta             | 1400 | 0.18         |
| W              | 2600 | 0.3          |

\*all values subject to wide variations depending upon material preparation.

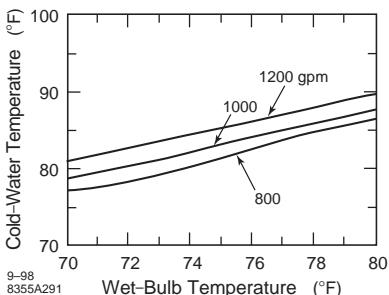


Figure 3: Cooling tower performance. Design for 85°F cold water, 95°F hot water, 78°F wet bulb temp., 1000 gallon per minute, gpm, cell.

Radiation shields are isolated reradiating surfaces interposed to reduce energy transfer. For the situations to which the above expression applies, i.e. all of the radiation leaving an enclosed surface falls on its enclosing surface, and for all surfaces having the same emissivity,  $N$  shields reduces the energy exchange calculated to  $P_{12}/(N + 1)$ .

**Evaporative cooling** The heat of transformation for water can be used directly for cooling, the heat of vaporization being 2260 kJ/kg and of fusion being 333 kJ/kg. For large installations, forced air evaporative cooling towers are usually employed. The low conductivity water coolant is circulated through an air-to-water heat exchanger over which is sprayed the water whose evaporation performs the desired cooling. The effectiveness of the tower depends upon the local wet bulb temperature. The common terms of reference are the approach ( $\Delta T$  between the cold water leaving the tower and the ambient wet bulb temp.) and the cooling range ( $\Delta T$  between the hot water into and cold water leaving the tower). Fig.3 [3] displays the performance of a typical commercial unit of 1.5 MW nominal capacity in US engineering units (see Sec.1.4 for conversion factors).

**Corrosion in water systems** Demineralized low conductivity water, 5 to 20 MΩ-cm is usually employed in cooling accelerator components and sub-systems, the higher value being used where direct or induced voltages tend to accelerate corrosion. A polishing system for continuously maintaining the resistivity is required. The pH of the circulating water should be maintained at about 8±1. The materials exposed to water should be chosen for electrolytic compatibility, e.g. all Cu or all Al. Stainless steel can be combined with copper if insulating flanges or hoses

are used for the interface. In all-copper systems, dissolved O<sub>2</sub> forms CuO in the warmest parts of the system, a very serious problem in Cu systems that run hot as the CuO flakes away, thinning the metal. Continuous purging of the system with N<sub>2</sub> is an effective counter measure as the N<sub>2</sub> displaces the O<sub>2</sub>.

## References

- [1] Reference Data for Radio Engineers, H.W. Sams Pub. (1970)
- [2] L.J. Giacoletto, Ed. Electronics Designer's Handbook, McGraw-Hill (1977)
- [3] E.A. Avallone, T. Baumeister, Marks Standard Handbook for Mechanical Engineers, McGraw-Hill, 10th ed. (1996)
- [4] R.H. Perry, D.M. Green, J.O. Maloney, Perry's Chemical Engineer's Handbook, McGraw-Hill (1984)
- [5] I.E. Idelchik, Handbook of Hydraulic Resistance, Hemisphere Pub., 2nd ed. (1986)
- [6] F.P. Incropera, D.P. DeWitt, Fundamentals of Heat and Mass Transfer, Wiley, 3rd ed. (1990)
- [7] W.H. McAdams, Heat Transmission, McGraw-Hill, 3rd ed. (1954)
- [8] H.S. Carslaw, J.C. Jaeger, Conduction of Heat in Solids, Oxford U., 2nd ed. (1959)
- [9] W.M. Kays, Convective Heat and Mass Transfer, McGraw-Hill (1966)
- [10] W.M. Kays, A.L. London, Compact Heat Exchangers, McGraw-Hill, 2nd ed. (1964)
- [11] Materials at Low Temperatures, R.P. Reed and A.F. Clark Eds., American Society for Metals, 44073 (1983)
- [12] S.W. Van Sciver, Helium Cryogenics, Plenum (1986)
- [13] G.E. Childs, L.J. Ericks, R. Powell, Thermal Conductivity of Solids At Room Temperature and Below, A Review and Compilation of the Literature, US National Bureau of Standards Monograph 131
- [14] L.S. Tong, Y.S. Tong, Boiling Heat Transfer and Two-Phase Flow, Taylor and Francis (1997)
- [15] Handbook of Single-Phase Convective Heat Transfer, S. Kaka, R. Shah, W. Aung, Eds, Wiley (1987)
- [16] Handbook of Heat Transfer Fundamentals, W.M. Rohsenow, J. P. Hartnett, E. Ganic, Eds, McGraw-Hill, 2nd ed. (1985)
- [17] Thermophysical Properties of Matter, 13 Volumes, Plenum (1970 - 77)
- [18] US National Institute for Standards and Technology, Standard Reference Data Program,

221/A320, Gaithersberg, MD 20899 • sr-data@nist.gov

The catalog of Standard Reference Data Products is available on the web at <http://www.nist.gov/srd/>

- [19] CRYOCOMP and METALPAK Computer Property Packages, Published by Cryodata, Inc., available on the web at <http://www.h tess.com/software.htm>

## 5.8 FABRICATION OF NIOBIUM RF STRUCTURES

*T. Hays, H. Padamsee, Cornell U.  
D. Proch, DESY*

(See Secs.7.3.9, 7.3.11 and [1] for cavity design, performance and construction.)

Cavity fabrication and preparation procedures discussed here are reviewed in [1, 2]. Niobium cavities are constructed from sheet niobium by deep drawing or spinning half cells, followed by trim machining and electron beam welding. The purity of Nb purchased is important, both in terms of dispersed impurities and inclusions from manufacturing steps, such as rolling. Inclusions on the rf surface play the role of normal conducting sites for thermal breakdown. Dissolved impurities serve as scattering sites for the electrons not condensed into Cooper pairs, to lower the thermal conductivity and thereby limit the maximum tolerable surface magnetic field. See Sec.7.3.9 for expected thermal conductivity and related residual resistivity ratio contributions due to dissolved impurities.

Among the metallic impurities, tantalum is the highest concentration (typically 500 ppm by weight). This impurity level is normally not harmful since tantalum is a substitutional impurity and does not substantially affect the electronic properties as do interstitial impurities. Tantalum could prove dangerous if it clusters to become a normal conducting spot, but this is very rare.

Among the light, interstitially dissolved impurities, oxygen is dominant. The other common interstitials are carbon, nitrogen and hydrogen. The electron scattering effectiveness of the various impurities are shown in Sec.7.3.9, Tab.5, in terms of the RRR of Nb for 1 ppm/wt of each impurity. The electron scattering contribution of the phonons is always present, so that the highest theoretical RRR for Nb is 35,000.

The interstitials are removed during the electron beam melting stages of the ingot [3]. Multiple melts and progressive improvements in the furnace chamber vacuum have led to RRR values of 300 which can be routinely supplied.

Niobium sheets can be ordered from several suppliers to meet key cavity specifications: RRR >300, with Ta content less than 500 wppm, yield strength greater than 50 MPa (N/mm<sup>2</sup>), tensile strength >100 MPa and percent elongation greater than 30%, uniform grain size ASTM 6 (~50 μm), thickness variations ±0.1 mm, planarity tolerance <0.5 mm. The sheets must be inspected for flatness, uniform grain size, near-complete re-crystallization, RRR value, and good surface quality, such as absence of scratches.

The Nb material should be free of defects (foreign inclusions, pits, or cracks and laminations) which will initiate a thermal breakdown. Such defects may be detected by quality control methods such as eddy current scanning [4] or squid scanning [5].

**Fabricating niobium parts** In deep drawing, a blank is pressed into the half-cell shape using a set of dies as depicted in Fig.1 for a 1.5 GHz cavity from 3 mm thick sheet Nb. These dies are machined out of 7075-T6 aluminum alloy because of its high yield strength, ease of machining, and low cost. The starting blank for deep drawing is a Nb disk cut from a sheet. It is bolted across the female die with a hold-down plate. With appropriate torque on the bolts, the outer edge of the Nb is constrained without tearing at the clamped edges. Clean motor oil is painted onto the Nb for lubrication. (Alternatively an elastomer sheet replaces the oil film.) For the 1.5 GHz cavity half-cell shown, about  $5 \cdot 10^5$  N are applied in a hydraulic press. To get the curvature required at the iris, the nose of the cup can be more precisely shaped by a further step of coining. In almost all cases when proper mechanical properties of the sheet are achieved, cavity parts can be deep drawn (or spun) to final shape without intermediate annealing because niobium has a low degree of work hardening. If the grain are too large (e.g. ASTM 4), or grain size non-uniform, an “orange peeling” effect occurs. If the material is incompletely recrystallized, it tears during deep drawing. Achieving good mechanical properties for high RRR Nb requires proper choice of annealing temperature and time. When making large half cells, (i.e. f <500 MHz), spinning is often

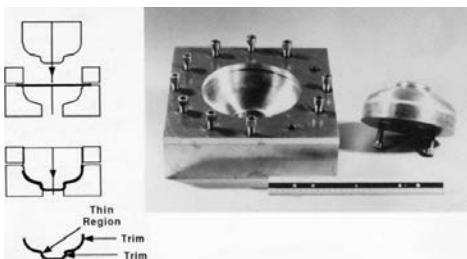


Figure 1: Deep drawing a Nb disk into a half cell.

preferred over drawing to eliminate the need for a high tonnage hydraulic press.

After forming, the half-cells are checked with a Co-ordinate Measuring Machine (CMM) for correct shape, and trimmed to the final size for electron-beam welding. Trimming is done either on a lathe or a CNC (computer numerical control) milling machine. It is important to maintain tight tolerances in weld zones ( $<50 \mu\text{m}$ ) for precise welds. Some practitioners use a step for registration of mating parts, but this can present a welding problem if there are small gaps. Parameters are now available (see below) for performing direct matching “butt” joints, especially at the equators. When trimming the formed parts, good cooling must be provided (e.g. Accu Lube<sup>TM</sup>).

Beam tubes are best purchased as drawn tubes. If the magnetic field at the beam-tube iris transition exceeds 20 mT at the design gradient, the beam tube should also be made from high RRR Nb. If tube material is not available the beam tube can be rolled from sheet and electron-beam welded. The input coupler and Higher Order Mode (HOM) coupler ports are part of the beam tube end-assemblies. A small section of the port is first extruded from the tube, which is then re-formed by passing a round plug through the tube. The remainder of the coupler port is electron beam welded on to the starting section to finish the cavity-end assemblies. Flanges for the beam tubes are machined from low RRR, reactor grade niobium because they are in regions of low magnetic field.

**Pre-cleaning** After degreasing in soap and water, the formed Nb parts are soaked in 50% sulfuric acid overnight to remove ferrous surface impurities. The Nb parts then receive a light chemical etch to remove 5 to 10  $\mu\text{m}$  from the surface. The etching agent is a buffered chemical polish (BCP) of 1 part hydrofluoric acid, 1 part nitric

## Ch.5: MECHANICAL CONSIDERATIONS

acid and 2 parts phosphoric acid (the 1:1:2 acid mix refers to volume out of the bottle (US). In the US, the concentrations in the bottles are 49% for HF, 69.5% for  $\text{HNO}_3$  and 85% for  $\text{H}_3\text{PO}_4$ .) The acid is cooled to 15 °C to avoid hydrogen pickup (see below). The etch rate is about  $1 \mu\text{m}/\text{min}$ , and should be locally calibrated with coupons. Parts should be soaked in clean water overnight, and carefully inspected for rust spots. If iron is found, the sulfuric acid treatment and inspection must be repeated.

**Electron Beam Welding** All parts are electron-beam welded together in a vacuum of better than  $2 \cdot 10^{-5}$  Torr to maintain RRR. The weld parameters are chosen to achieve full penetration welds with a smooth underbead by using a defocused electron beam (Fig.2). It is important NOT to use a focused beam to avoid weld spatter, resulting in myriad weakly attached Nb beads and balls. The vapor column from a focused beam may also leave voids in the molten Nb weld seam. One technique to produce a well defined and reproducibly defocused beam is to raster a focused beam to cover a pattern in the shape of a rhombus, commonly referred to as the “rhombic raster weld” [6], Fig.3. Typical parameters to achieve full penetration in 1.6 mm thick Nb are  $V = 50 \text{ kV}$ ,  $I = 37 \text{ mA}$  at a 46 cm/min weld speed. For the 2.6 mm thick, butt weld the parameters are  $I = 56 \text{ mA}$  and speed of 25 cm/min.

The joint at the cell’s equator is a region of high surface current and will show strong heating if the weld is done poorly. For the thicker material (2.6 mm) there is a small under-cut to the underside of the weld (Fig.4) [7] which is filled in by the normal molten bead drop through, and leaves a near-flat surface on the rf side. The edges of the under-cut on the rf side should be rounded in case the weld does not fully consume the step area. Weld parameters should give reproducible shrinkage, minimal distortion and tolerable margin for the beam current needed to produce a good weld.

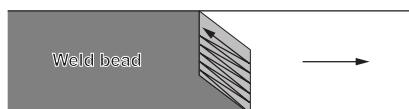


Figure 2: Rhombic raster pattern for the beam during an electron beam weld. This rastering produces a well defined beam needed for trouble free welding of Nb. The rastering spreads the beam over a large area.



Figure 3: The smooth underbead of a full penetration e-beam weld in 1.5 mm Nb.

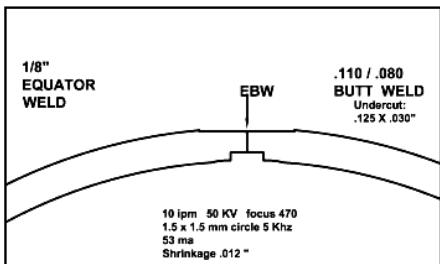


Figure 4: Weld geometry for smooth defocused weld on thick material.

**Weld Sequence** The iris welds between half-cells, and the end components with beam tubes and coupler ports are welded first. The equators are then electron-beam welded together to complete the cavity. Sufficient time should be allowed between welds for parts to cool down to avoid the likelihood of holes due to overheating.

All welds are inspected for complete, smooth under-bead, flat on the inside, and no weld spatter using a borescope if necessary. If a weld is rough, it may be necessary to perform an extra step of grinding the weld to ensure a smooth inner surface. Grinding is done with a ScotchBrite<sup>TM</sup> abrasive wheel. After the grinding, the cavity pieces are cleaned by etching with 1:1:2 BCP to remove about 20  $\mu\text{m}$  from the surface. Tumbling or Centrifugal Barrel Polishing (CBP) can also be used for smoothening rough welds. This procedure provides a fairly uniform surface by removing imperfections such as roughness at welds, pits, and mild scratches remaining from the starting sheet material. The rate of material removal is highly dependent on the tumbling medium and rotation speeds. A light step of buffered chemical polishing (BCP), usually about 50  $\mu\text{m}$ , removes the tumbling abrasive embedded in the surface.

The inside surface is inspected optically, especially the weld quality. Special inspection tools have been developed [8]. Mechanical measurements ensure straightness and correct dimensions. The cell-to-cell field profile is checked and adjusted. The usual goal is 98% field flatness.



Figure 5: Closed acid chemistry system at Cornell. The acid is stored in a chilled reservoir before being admitted to the cavity. Valves permit the cavity to be safely filled with acid and drained and then filled with ultrapure water.

**Chemical treatment** A clean rf surface is finally achieved by chemically etching away a surface layer 100  $\mu\text{m}$  to remove the surface damage layer due to sheet and cell manufacturing and vapor deposited from welding. BCP is used for applications requiring surface fields less than 80 mT, corresponding roughly to  $E_{acc} < 20 \text{ MV/m}$  for 1.3 GHz elliptical cavities. Electropolishing (EP) is needed to provide a smooth surface for applications demanding  $E_{acc} > 20 \text{ MV/m}$ .

For BCP, the cavity is filled with acid using a “closed” system. Valves can be remotely opened and closed remotely to fill the cavity with cooled acid, drain the acid, and rinse the cavity with ultrapure water. The rinsed water is sent to a neutralization tank. See Fig.5. Experience has shown that hydrogen contamination can be avoided by keeping the acid temperature  $< 15^\circ\text{C}$ .

For EP surface preparation [9], the cavity is the anode (+) in an electrolytic cell and the cathode (-) is made from pure aluminum (1100 series). The electrolyte is a mixture of hydrofluoric and sulfuric acid in a volume ratio of 1:9, using typical US commercial strengths HF (40%) and H<sub>2</sub>SO<sub>4</sub> (98%). The associated molar concentrations are 2.29 mol/L of HF and 16.08 mol/L of H<sub>2</sub>SO<sub>4</sub>. The nominal HF content for the 9:1 mixture is  $\sim 100 \text{ cm}^3/\text{L}$ . Under optimum conditions of voltage, temperature and stirring, the current shows oscillations (0.5 - 5 Hz at 12 V and 30°C). The best current density is between 30 - 100 mA/cm. Figure 6 shows the schematic [10] for horizontal EP of a single cell. Rotation of the cavity polishes the

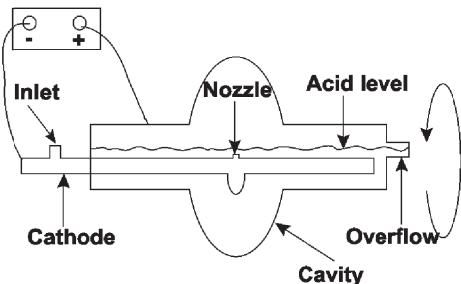


Figure 6: Schematic for continuous EP [12].

entire surface, while continuous flow of the electrolyte between the cavity and an acid reservoir provides the acid circulation. A membrane pump drives the acid mixture through the cooled barrel, and through a  $1\text{ }\mu\text{m}$  pore filter before it reaches the inlet of the hollow cathode. Hydrogen evolves at the (negative) cathode and rises to the electrolyte surface to form a thick foam of bubbles. About 2 L/min of H gas will be evolved at 300 A. A perforated teflon coaxial tube (or porous teflon cloth) surrounds the hollow aluminum cathode to inhibit the evolving hydrogen gas from mixing with the electrolyte and reaching the niobium surface. Dry nitrogen gas floods the space above the electrolyte to prevent water vapor from reacting with the hygroscopic  $\text{H}_2\text{SO}_4$ . Maintaining the relative composition of HF and  $\text{H}_2\text{SO}_4$  is important for best polishing, and to avoid synthesis of solid sulfur particles. The conditions for best surface roughness and brightness correspond to an HF content of  $>60\text{ cm}^3/\text{L}$ . An indicator of electrolyte aging is the Nb content in solution. After the electrolyte gains 5-7 g/L Nb, the surface starts to degrade in brightness and there is increased S deposit.

The average material removal is determined by the integrated charge, ultrasonic thickness measurement, or by weighing the cavity accurately before and after EP. The ideal removal rate is  $0.67\text{ }\mu\text{m}/\text{minute}$  for  $50\text{ mA/cm}^2$  current density. The actual removal rate is about  $0.4\text{ }\mu\text{m}/\text{minute}$  since only a fraction of the cavity area is in contact with the acid. Due to the proximity of the cathode to the irises, the etch rate at the equator is about 60 % the average removal rate.

A less expensive EP method with the cavity in the vertical orientation is under development. Acid fills only the cavity volume, so a smaller quantity is required. Stirring is by internal

paddles, and cooling is by water flowing directly over the outside of the cells, so that there is better control of acid flow and temperature.

**Rinsing** After BCP the cavity is thoroughly rinsed with high purity water and transported to a Class 10 - Class 100 clean room. After EP, a cavity is ultrasonically rinsed in a dilute surfactant<sup>1</sup> (Micro-90) for several hours to remove any S deposits, followed by rinsing with water. An alternative approach to dissolve S is to rinse the inside surface for one hour in ethanol, rinse with water and transport to the Class 100 clean room.

In the clean room the surface is given a high pressure (100 bar) rinse (HPR) with high purity water jets for many hours (typically 2 hours per 1.3 GHz cell). The main goal is to scrub the chemical residues and particulate contaminants which may cause field emission or thermal breakdown [11]. It is important to avoid drying between final water rinse after chemistry and the start of the first HPR. Particles adhere more strongly to the surface as the water evaporates. Typically HPR systems work with a water flow rate between 5 and 20 l/min and a pressure between 80 - 150 bar, which allows removal of particles larger than a few micrometer. The final particle filter (pore size  $\leq 0.2\text{ }\mu\text{m}$ ) is placed as close to the nozzle as possible with no moving parts (i.e. valves) or dead ends between filter and nozzle. It is important to monitor the erosion of the nozzles, especially if the shower head is made from stainless steel. The preferred nozzle material is sapphire. For one system (at DESY) the water exits through 8 nozzles in the head at the top of the rinsing cane, which moves up and down inside the cavity while the cavity rotates. During a 2 hour rinse up to 2000 liters of water are sprayed onto the surface.

Water quality deserves close attention. The standard practice to monitor the pure water supply systems is to measure resistivity, particle count, bacteria, total organic content (TOC), and residues. Particle counters monitor particles in the sub- $\mu\text{m}$  range. The resistivity of the water should be close to theoretically pure ( $>18\text{ M-cm}$ ) and the water inlet should be filtered to eliminate particles  $>0.3\text{ }\mu\text{m}$ .

After HPR, the cavity dries in the clean room for 1 - 2 days. Clean pumping and heat lamps can be used to speed up the drying process.

<sup>1</sup>surface active agent

**Hydrogen Degassing** A potential danger during chemical treatment is a low Q caused by hydrogen contamination (“Q-disease”) [12]. This can be avoided for BCP by keeping the acid temperature below 15° C, but is more difficult to avoid during EP. Dissolved H is removed by heating in a vacuum furnace at 800°C for 2 hours, or 600° C for 10 hours. This reduces the hydrogen concentration to a few atomic ppm in the bulk and  $\leq 1$  atomic % in the surface layer. The lower temperature annealing avoids the danger of grain growth and yield strength drop, especially for material RRR  $\geq 300$ .

After H-degassing the cell-to-cell field profile needs to be measured and adjusted by mechanical tuning due to non-uniform material removal during the bulk chemistry and possible deformation during heat treatment. A final short EP removes 25  $\mu\text{m}$  along with dust and contaminants deposited during the furnace treatment or field profile flattening. The rinsing, HPR and drying described above is repeated to ensure a final clean surface for the performance qualifying test in the vertical dewar.

**Clean assembly** A second HPR and drying takes place after assembly of necessary flanges, and the field monitor probe. Several final assemblies follow: attaching flanges, installing input couplers, or assembling cavity strings for cryomodule preparation. Great care must be exercised to avoid recontamination during the subsequent cavity handling, component assembly, installation and operation of the accelerator modules. Handling and assembly time of an open cavity should be kept as short as possible. Movement of personnel should be slow and monitored by particle counters to avoid kicking up dust, especially from the floor. Clean Room staff must be well trained for the high level of cleanliness necessary.

All components, including fastening hardware (screws and nuts), are pre-cleaned in an ultrasonic-bath and rinsed with pure water. After drying in the clean room the residual particle contamination is controlled before installation of components to a cavity. All components are blown carefully by ionized and particle filtered (0.02  $\mu\text{m}$ ) nitrogen gas. Satisfactory attenuation of the number of particles blown off is monitored by an air particle counter. Care must be taken to ensure that the vacuum system is thoroughly clean and dust free. The cavity itself is evacuated slowly

<0.1 torr per sec to avoid turbulent flow, and reduce the risk of contaminants from the vacuum system reaching the cavity surface. It is essential to maintain dust free conditions while attaching input and output couplers to the cavity.

**Mild Baking** To reach the highest fields, an electropolished cavity needs to be baked at 120° C for 48 hours to remove the high-field Q-drop which sets in above  $E_{acc} > 20$  MV/m [10]. This bake is normally carried out with the inside of the cavity in a good vacuum (about  $10^{-8}$  torr).

**Future Developments** We have restricted our discussion to methods which have been successfully used to fabricate cavities for on-going projects, and left out potential developments that are likely to take place in the coming years.

Large grain niobium cut directly from a large grain ingot is under development to lower the cost of the sheet and improve sheet purity. Another cost benefit of large grain Nb is that BCP can be used instead of the more expensive EP. Cavities made from large grain prepared by BCP reach fields  $> 30$  MV/m after the mild bake without a Q-slope at high fields [13].

The sheet metal forming, machining and welding method is time-consuming and labor intensive. Therefore development continues on several alternate cavity fabrication methods such as hydroforming [14] and spinning. Research continues into thin film deposition by sputtering and other techniques but suffers from the problem of Q-slope which starts at low fields. A 1300° C post-purification heat treatment procedure is helpful to raise the RRR by heat treatment at 1300° C with titanium. Since high fields can generally be reached by EP there is less motivation for post purification, especially since the high temperature treatment drops the niobium yield strength substantially, making the cavities vulnerable to plastic deformation and frequency detuning. Yield strength can decrease to 40 MPa or even lower.

## References

- [1] H. Padamsee, J. Knobloch, T. Hays, RF Superconductivity for Accelerators, Wiley (1998), 2nd Edition, Wiley-VCH, 2008
- [2] H. Padamsee, RF Superconductivity, Science, Technology and Applications
- [3] H. Padamsee, Proc. 6th Workshop on RF Superconductivity (1994) p.515

- [4] W. Singer, Proc. 8th Workshop on RF Superconductivity, p. 850 (1998).
- [5] W. Singer, et al, A. Farr, et al, Proc. 12th Workshop on RF Superconductivity, Ithaca, NY, USA, paper TuP49 (2005).
- [6] J. Kirchgessner, Proc. 3rd Workshop on RF Superconductivity (1988) ANL-PHY-88-1, p.533
- [7] J. Sears, Proc. 12th Workshop on RF Superconductivity, Ithaca, NY, USA, p.481 (2005).
- [8] Y. Iwashita, Proceedings of EPAC 2008, Genoa, p. 1956.
- [9] K. Saito et al., Proc. 4th Workshop on RF Superconductivity, KEK, Tsukuba, Japan, p.18 (1989).
- [10] L. Lilje et al, Nucl. Instr. Meth. Phys. Res. A, 516, p. 213 (2004).
- [11] P.Kneisel et al., Proc. of the 6th Workshop on RF Superconductivity, Newport News, VA USA, p. 628 (1993).
- [12] B. Bonin, R. Roeth, Proc. 5th Workshop on RF Superconductivity, p.210, 426, DESY -M-92-01 (1991)
- [13] P. Kneisel, et al, Proc. 12th Workshop on RF Superconductivity, Ithaca, NY, USA, paper p.134 (2005).
- [14] W. Singer, Proc. 12th Workshop on RF Superconductivity, Ithaca, p.143, NY, USA (2005).

## 5.9 REFRIGERATION SYSTEMS

*C. Rode, R. Ganni, TJNAF*

This section will concentrate on liquid helium (LHe) systems. There are three parts to these systems: (i) the production - the refrigerator, (ii) storage and utilities, and (iii) the distribution - the transfer lines and valve boxes. The transfer lines are often thought of as a minor item; this is not true; several commercially fabricated and installed lines have cost almost as much as the refrigerators. Tab.1 lists the world's largest He systems and major parameters.

### 5.9.1 Refrigerators

We loosely call all cryogenic plants, refrigerators; more accurately there are three types: (i) *Liquefier*: Liquefaction is that liquid flow-rate by which the cold end supply flow exceeds the return flow; an example is the filling of a dewar. (ii) *Refrigerator*: Refrigeration is cooling power produced with the cold end supply and return flow equal. (iii) *Satellite*: The return flow is larger than the supply; these units consume liquid to produce refrigeration ("economizer" cycle). These can be

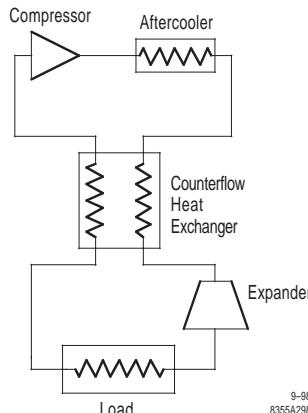


Figure 1: Refrigerator building blocks.

thought of as flow amplifiers with a gain of about 12; an example of this is the Tevatron Cryogenic System which uses 24 units to provide distributed refrigeration.

All plants can be operated in all three modes but they become increasingly inefficient the further they operate from the design point.

**Power and efficiency** The efficiency  $\eta$  of a refrigerator is given by the ratio of Carnot power to the room temperature power and varies from 10% to 30% for the 100 to 20,000 W plants. For plants precooled by liquid nitrogen ("Brayton" cycle) the power to produce the LN<sub>2</sub> must be included. The power required to produce refrigeration is given by

$$P_h = \frac{1}{\eta} \frac{T_h - T_c}{T_c} P_c = \frac{1}{\eta} \frac{T_h - T_c}{T_c} \Delta h_L \dot{m} \quad (1)$$

and for liquefaction by

$$P_h = \frac{\dot{m}}{\eta} [T_h (S_h - S_c) - (h_h - h_c)] \quad (2)$$

where  $P_h$  and  $P_c$  are the hot and cold powers [W];  $T_h$  and  $T_c$  are hot and cold temperatures [K];  $\Delta h_L$  is the latent heat of evaporation [J/g];  $\dot{m}$  is the net flow [g/s];  $S_h$ ,  $S_c$ ,  $h_h$ , and  $h_c$  are the hot and cold entropy [J/g-K] and enthalpy [J/g]. For a cold end of 4.5 K and room temp. of 300 K, these equations simplify to

$$\begin{aligned} P_h &= 65.67 P_c / \eta \quad (\text{refrigeration}) \\ P_h &= 6580 \dot{m} / \eta \quad (\text{liquefaction}) \end{aligned} \quad (3)$$

Table 1 Large helium refrigerators and liquefiers

|       |                    | Temperature [K] | Capacity [kW]   | Units      | Status                           |
|-------|--------------------|-----------------|-----------------|------------|----------------------------------|
| 1     | LHC                | ~4.5            | 4x18            | 4+4        | Operational                      |
| 2     | LEP                | ~4.5            | 2x6<br>4x18     | 2<br>4     | Operational<br>Converted for LHC |
| 3     | RHIC               | 4.5 (55)        | 30 (60)         | 1          |                                  |
| 4     | TEVATRON           | 4.5<br>4.6      | 30<br>10x0.6    | 2+29<br>10 | Operational                      |
| 5     | HERA               | 4.35 (60)       | 3x6.3 (3x13)    | 3          | Operational                      |
| 6     | EXXON              | ~4.4            | 2x 2600 l/hr    | 2          | Operational                      |
| 7     | MFTF               | 4.35            | 10+3.3          | 2          | Decommissioned                   |
| 8     | CEBAF              | 2.0 (45)<br>4.5 | 4.8 (12)<br>1.5 | 1<br>1     | Operational                      |
| 9     | CITIES<br>SERVICES | ~4.4            | 2400 l hr       | 1          | Operational                      |
| 10    | TRISTAN            | 4.4             | 4.5/6.6         | 1          | Decommissioned                   |
| 11-30 | Misc. Systems      | ~4.5            | 1...4           | 1          | Operational                      |

Note that 1 g/s of liquefaction is thermodynamically equivalent to 100.2 W of refrigeration. For small 4.5 K plants assuming  $P_h/P_c = 500$  W/W is a reasonable design, while large DESY and CERN plants do better than 300 W/W. The CEBAF 2 K system is 1000 W/W and the LHC 1.8 K system is 950 W/W. For 3 bar 88 K LN<sub>2</sub> and room temperature of 300 K (which will have a Carnot efficiency in the order of 35%) these equations simplify to: ( $\eta=35\%$ )

$$\begin{aligned} P_h &= 6.883 P_c \quad (\text{refrigeration}) \\ P_h &= 1754 \dot{m} \quad (\text{liquefaction}) \end{aligned} \quad (4)$$

**Building blocks** The refrigerator is comprised of five building blocks as shown in Fig.1; this is the simplest cryogenic refrigerator (Reverse Brayton Cycle) and is used to produce 50 K refrigeration. The large He plants are obtained by stacking these building blocks for efficiency; the CEBAF plant is shown as an example in Fig.2.

**Compressors Warm Compressors:** The standard compressor today is the oil injected screw compressor. It was chosen for high reliability (50,000 h mean time between major failures). It

replaced the much higher efficiency but lower reliability reciprocating piston compressor.

The oil injection provides, in addition to lubrication, the rotor sealing and extraction of the compression heat, thus approximating isothermal cooling. Typically 2/3 of the inefficiency of a 4.5 K system is due to this component [1]. For larger units the isothermal efficiency drops from 55 to 45% as the compression ratio increases from 3 to 7. Isothermal efficiency is defined as

$$\eta_{\text{isoth}} = \frac{P_{\text{isoth}}}{P_h}, P_{\text{isoth}} = 10^5 V_1 p_1 \ln \frac{p_2}{p_1} \quad (5)$$

where  $V_1$  is the inlet volumetric flow rate (m<sup>3</sup>/sec),  $p_1$  and  $p_2$  are the inlet and outlet pressures (bar).

**Cold compressors, pumps and circulators:** For the very small systems with a volumetric flow less than 2 l/s reciprocating compressors are often used. They have one major advantage over turbo machinery: they can pump to high pressures, 5-20 bar in a single stage.

Turbo machinery has become the standard due to its much higher reliability and capacity; it

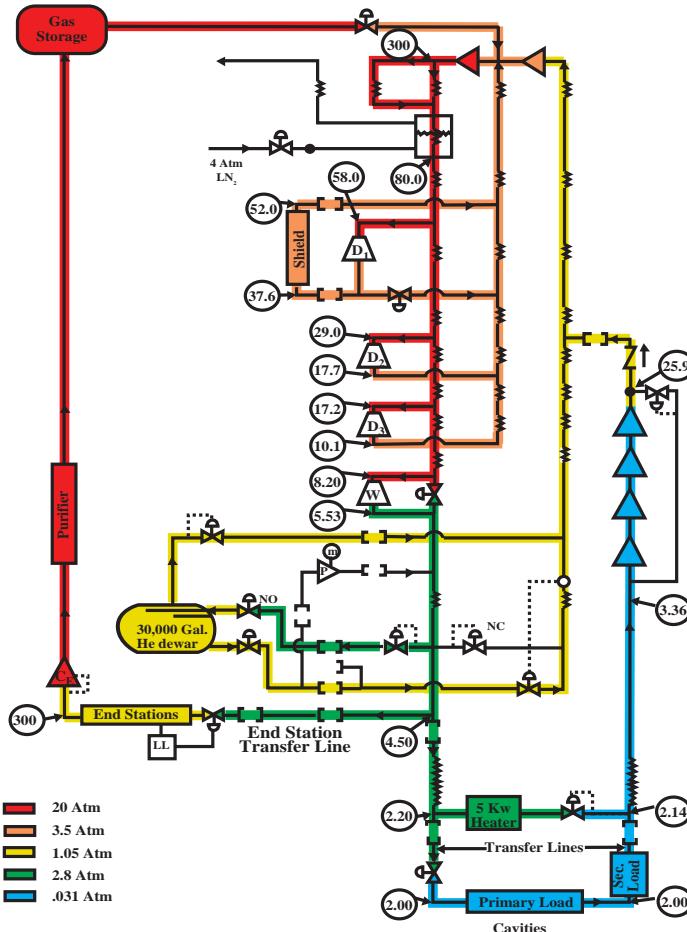


Figure 2: CEBAF Process cycle.

always has its efficiency defined in terms of adiabatic rather than isothermal efficiency,

$$\eta_{\text{adiabatic}} = P_{\text{adiabatic}}/P_c \quad (6)$$

$$P_{\text{adiabatic}} = 10^5 V_1 p_1 \left( \frac{\gamma}{\gamma - 1} \right) \left[ \left( \frac{p_2}{p_1} \right)^{\frac{\gamma-1}{\gamma}} - 1 \right]$$

where  $\gamma = C_p/C_v$ , the heat capacity ratio. It typically has an efficiency  $\sim 60\text{-}70\%$ .  $\gamma$  for He above 30 K is equal to 1.667.

Cold Compressors typically provide a compression ratio of 2 to 3 per stage while Pumps and Circulators provide a 1 bar rise per stage.

After cooler The oil injected screw compressor actually uses two heat exchangers to remove the

heat of compression. The first is the traditional after-cooler that re-cools the gas back to room temperature. 90% of the heat is removed by the oil injection; therefore an oil heat exchanger is the primary cooling element.

Counterflow heat exchanger Counterflow heat exchangers are required for efficient cycles as well as to achieve temperatures less than 180 K with He or 150 K with Freon and vacuum pumping. The critical parameter in the design of a cycle is the "pinch" of the cooling curves or minimum temperature difference between the supply and return streams. The pinch determines the inefficiency due to (i) the difference in enthalpy of the high pressure and low pressure streams, and

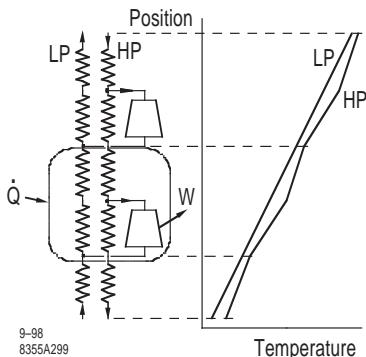


Figure 3: Heat exchanger cooling curves.

(ii) the fact that heat exchangers are finite. Fig.3 shows the central section of a typical refrigerator as well as the supply temperature plotted with respect to the return.

For a refrigerator with LN<sub>2</sub> precooling the first pinch is at 80 K and is less than two degrees. The temperatures of the next pinches are dependent on the number of turbines; for the CE-BAF design they were at 36 K (1.0 K), 18 K (0.6 K), 10.7 K (0.4 K), 4.5 K (0.10 K), and 2.2 K (0.06 K).

In order to analyse the cycle, one can draw a circle around a section and then apply the First Law of Thermodynamics; the sum of each flow (+ for inward and - for outward) multiplied by its enthalpy is equal to the refrigeration minus the heat leak,

$$\sum_i F_i H_i = W - \dot{Q} \quad (7)$$

By stacking 3-10 circles and using a simple calculator, one can design all the refrigerators over the size of 100 to 20,000 W. A computer is required for the last 10% of optimization; as well as the second phase, comparing the actual system with the design.

The heat transfer in an exchanger is given by

$$\dot{Q} = UA\Delta T_m \quad (8)$$

where  $U$  is the overall coefficient of heat transfer and  $A$  is the heat transfer area upon which  $U$  is based.  $\Delta T_m$ , the log-mean temperature differential is the effective temperature differential,

$$\Delta T_m = \frac{\Delta T_h - \Delta T_c}{\ln(\Delta T_h / \Delta T_c)} \quad (9)$$

where  $\Delta T_h$  and  $\Delta T_c$  are the hot and cold temperature differences between the supply and return flows. Typical values for  $U$  [2] are  $\sim 30-300$  W/(m<sup>2</sup> K). The value depends on the allowable pressure drop, which in turn depends on the pressure level of the stream.

Expander The efficiency of an expander is

$$\eta_{\text{adiabatic}} = \Delta h / \Delta h_{\text{adiabatic}} \quad (10)$$

The efficiency will vary from slightly negative to 80% for the larger units. The negative efficiency is produced by the static heat load. There are three types of expanders: a) valve, b) reciprocating, and c) rotary.

Valve: The zero static heat load valve is the definition of an expander with zero efficiency. The Joule-Thompson valve is the common name for an expansion valve.

Reciprocating: For inlet volumetric flow  $< 1$  l/s reciprocating hardware is used; powers vary from about 100 to 1000 W per piston and for ideal gases is given by the relationship

$$W = 10^5 \eta V_1 p_1 \left( \frac{\gamma}{\gamma - 1} \right) \left[ 1 - \left( \frac{p_2}{p_1} \right)^{\frac{\gamma-1}{\gamma}} \right] \quad (11)$$

where  $W$  is the refrigeration in watt which is independent of the temperature and  $V_1$  is proportional to the speed.

Rotary: For inlet volumetric flow  $> 1$  L/s turbo expanders are used; their power varies from less than 1 kW to 50 kW for He systems, to many hundreds of kW for N<sub>2</sub> systems. For He systems there is significant advantage in efficiency to have more smaller expanders in order to keep the cooling curves closer together. The  $V_1$  in Eq.(11) is not constant but is given by  $V_1 = C_\nu \sqrt{T_1}$  where  $C_\nu$  is the flow coefficient [m<sup>3</sup>/(s K<sup>1/2</sup>)] which is a constant proportional to the turbine nozzle area. Note the  $\sqrt{T_1}$  dependence; the power is significantly higher for cooldown mode vs. steady state. This impacts the design of the brake wheel and exchanger as well as bearings.

Load The loads are functionally just heat exchangers and fall into four types: (i) Pool boil - CEBAF, LEP, TESLA: A large tank of liquid; (ii) Two Phase Flow - LHC [3]: An isothermal cooling of a liquid/gas mixture; this system requires compensation for inclined machines such as weir or active flow control; (iii) Subcooled or Super-critical Forced Flow - RHIC, SSC: Temperature

increases monotonically through the magnets until it reaches a re-cooler; (iv) Counter Flow Heat Exchange - Tevatron, HERA: Combines the advantages and disadvantages of the above two. The magnet coils are in the supply forced flow; at the end of the string of magnets is where the approximately isothermal cooling is provided by heat exchange with the return 2-phase liquid.

**Margin** The refrigerator capacity must be over sized to facilitate cooldown, for control stability, and for capacity degradation due to contamination etc.; as well as for uncertainty in component heat loads. The minimum margin is 25%; for well understood loads the Fermilab/ Jefferson Lab design rule is 50% margin for refrigeration and 100% for liquefaction.

Another Fermilab / Jefferson Lab design rule is for 100% margin in the transfer line flow capacity. While it is relatively easy to increase refrigeration capacity, it is almost impossible to change the transfer lines.

### 5.9.2 Storage and Utilities

The household air conditioner stores the system inventory in the compressor and piping; this is not practical for larger systems. This is especially true for cryogens with their liquid to room temperature expansion ratio of 800.

**Gas storage** While clearly the simplest form of storage, it is also the most expensive, costing more than five times the value of its content. For small systems the gas storage is sized to hold the entire inventory. Large systems often elect to store only 10-20% of the inventory as gas.

In the USA the propane tank due to its mass production is the standard storage vessel: at the National Lab's one sees rows of 30,000 gal. tanks. It is rated at 250 psi gal. which makes it convenient to store 2000 liquid liters equivalent of He.

**Liquid storage** Liquid storage costs about the value of its content, and therefore is the preferred storage for the large systems. The pressure rating on small dewars is only 10 psi gal.. While often argued that it permits faster restart, it also makes restarting much more complicated. A dewar requires a minimum of three connections: (i) supply (2 phase or equivalent), (ii) gas return, and (iii) liquid out.

**Purification and dehydration** The key to a reliable cryogenic system is contamination control/prevention. This has several phases; the Fermilab

/ JeffersonLab procedure starts with dehydration of the system to <0.1 ppm H<sub>2</sub>O followed by purification to <1.0 ppm N<sub>2</sub>. Either single pass nitrogen gas from LN<sub>2</sub> boil off or helium circulation with desiccant beds can be used for this purpose. If one has an all metal system including the load, and if it is completely leak tight, this is adequate and one can cool down.

Air leaks cause problems in two ways: first one has reverse diffusion of air into the He and secondly one buys contamination in the makeup He. Fermilab / Jefferson Lab buy >99.9975% pure (<5 ppm Ne) and then repurify it with stand alone 80 K charcoal bed purifiers before adding it into the system.

At both Labs the main refrigerators have dual 80 K adsorbers for O<sub>2</sub> and N<sub>2</sub>, permitting operating on one while the other is being regenerated. The CEBAF main refrigerator also has a single 20 K adsorber for Ne and H<sub>2</sub> (with a bypass valve used during regeneration). The CEBAF unit has been operating continuously for over six years without a warm-up of the cold box.

### 5.9.3 Transfer Lines

Except for very small systems, the refrigerator is not adjacent to the load and therefore they need to be interconnected by vacuum jacketed transfer lines. For extremely vibration sensitive loads, e.g., SRF cavities, one requires a physical separation to permit earth damping. The transfer lines have three areas of overlapping design requirements: (i) stability, (ii) connections, and (iii) heat loads.

**Stability** He systems below 10 K are inherently unstable due to the large density ratios over very small changes in temperature. Fermilab/ Jefferson Lab have studied this problem and have found adequate stability regions to permit all required operations but are unable to generate stability equations. The following is an approximate prioritized list of stability improving factors for lines with  $L/D > 10^4$ :

- (i) Eliminate all potential Thermo-Acoustic Oscillators;
- (ii) Higher operating pressure for supply lines ( $P > 2.6$  bar);
- (iii) Lower heat leak per unit length ( $< 0.005$  W/m);
- (iv) Higher pressure drop per unit length ( $\Delta p/p > 10^{-5}/m$ );
- (v) Temperature hot or cold (density  $\gg 60$  or  $\ll 60$  g/cm<sup>3</sup>).

Thermo-Acoustic Oscillators and other heat oscillations can easily overload any refrigerator. Static warm to cold transitions have been studied in detail [4]; can be unstable over a range of 1 to 150 mm if the temperature ratio is > 6. Fermilab determined that flowing lines oscillated as well. The standard solution is to lower the resonance frequency by adding gas volumes. The Fermilab / Jefferson Lab design rule is to use either a cold valve or a 20 K check valve on all lines larger than 0.7 mm. It must be noted that refrigerator manufacturers are able to ignore this problem and almost always are successful.

Jefferson Lab violated this rule for pressure taps and used 3.0 mm ID due to the continuous problem of 0.7 mm lines plugging with contamination. For 100 mm 2 K vacuum pumping lines (0.03 bar) the Fermilab / Jefferson Lab design rule cannot be used due to the associated pressure drops. Therefore an up-down-up three legged transition should be used; it has always been successful to date. An emergency fix is to inject 300 K gas to reduce the density; a temperature ratio of 100 appears to be stable for 0.03 bar.

An irritating aspect of heat oscillations is that they cannot be predicted; one can only say there is a potential for an oscillation. An example of this is the Fermilab 1/2" bellows control valve of which 300 are successfully in operation. The identical components at Jefferson Lab had a factor of 2000 reduction in MTBF; this was due to the environment in which they were installed. The Jefferson Lab systems have a much larger number of valves and bayonets per unit length, and therefore increased heat leak per unit length. Therefore the system is fundamentally less stable and components are more likely to oscillate.

Two phase flow The design must not be based on froth flow but must be capable of operation with perfect phase separation [5]. This implies that the minimum drive pressure differential must be greater than the sum of all the upward flow gravitational heads with no credit for the down legs,

$$\Delta p_{\min} = g\rho \sum L_{\text{up}} \quad (1)$$

where  $g$  is the gravitational intensity (1019.7  $\text{g}/\text{cm}^2/\text{bar}$ ),  $\rho$  is the liquid density ( $\text{g}/\text{cm}^3$ ),  $L_{\text{up}}$  is the liquid head (cm).

Supercritical flow While the He critical point is 2.27 bar, the minimum operating point must be 2.6 bar; below this pressure the properties of

supercritical and two phase are about the same. This can be understood by plotting the density as a function of temperature for various pressures above critical. Between 2.6 and 2.8, one gains another factor of five in stability. Fermilab / JeffersonLab operating points vary between 2.8 and 3.0 bar.

Subcooled liquid Subcooled Liquid is very stable due to its ability to adsorb heat without a significant density change. The only issue is cooldown; one must use one of the above two modes to cooldown.

**Connections** There are two approaches for connecting transfer lines to refrigerators and loads. The first and the simplest is to weld them directly to the hardware. Normally vacuum breaks are used to isolate subsystems in order to facilitate installation, troubleshooting, and repair. The prime example of this configuration is the HERA system.

The second is the Fermilab / Jefferson Lab bayonetted U-tube and L-tube concept. The bayonets have been designed to permit interconnections under operating cryogenic conditions. They consist of a labyrinth seal at the cold end and chevron seals and valve at the warm end [6].

**Heat loads** The transfer line heat loads can be a significant fraction of the total loads. This is particularly true of the shields for lines that run the full length of the accelerator; 30 to 40% of the total is not unreasonable.

Shields are normally used in lines  $\gtrsim 10$  m to intercept the radiative load as well as to a lesser degree the conductive load. The heat loads for the shielded lines (shielded with  $\text{LN}_2$  or low temperature helium) are considerably lower (up to an order of magnitude) than the unshielded values. For 4 K systems the shields are cooled to 60 K with He or 80-90 K with  $\text{LN}_2$ . For 2 K systems, shields below 50 K are appropriate.

Super insulation From room temperature to the shield or outer circuit 80-90% of the load is radiative. Multilayer aluminised mylar with spacers is commonly used to reduce the radiative heat load [7]. Inside the shield the radiative load is dominated by cracks in the shield. The radiative heat transfer per unit area is

$$\dot{Q} = \sigma_{SB}\epsilon(T_h^4 - T_c^4) \quad (2)$$

where  $\sigma_{SB}$  is the Stefan-Boltzmann constant and  $\epsilon$  is the emissivity (Sec.5.7).

For all practical purposes 30 layers of mylar with good vacuum is all it requires to reduce the radiative load significantly. Sometimes 60 layers are used to safeguard against operating the system under poor vacuum conditions. The nominal heat load from room temperature to the first shield with 30 layers of mylar super insulation is  $\sim 1.0$  W/m<sup>2</sup>. The transfer lines [8] at HERA achieved 0.6 W/m<sup>2</sup>. The LHC ring line [9] achieved 3 W/m on the thermal shield, corresponding to about 1 W/m<sup>2</sup> the 30 layers of superinsulation, and 0.2 W/m on the inner piping.

Supports As lines become larger, supports become increasingly important. This is due to the fact that the weight of the internal lines is growing at least with (diameter)<sup>2</sup>. Supports come in many styles and shapes but normally are fabricated from G10 or G11. The support design needs to take into account the weight, relative motion due to thermal expansion resulting from cooldown and warm-up, and anchor points of the lines with low heat leak to the process. Improperly designed supports become thermal shorts and when they are close to seals and O-rings will cause leaks. To a large extent the mechanical integrity of the vacuum jacket lines depend on the support design. The CEBAF transfer line system report[10] explains some of the parameters used in the design.

## References

- [1] B.O. Ziegler, Advances in Cryogenic Engineering, Plenum(1986) Vol.31, p.693
- [2] R.F. Barron, Cryogenic Systems, 2nd ed., Oxford (1985) p.109
- [3] Ph. Lebrun, IEEE Transactions on Applied Superconductivity, Vol 10, No 1 (March 2000), p. 1500
- [4] N. Rott, J. Appl. Math. & Phys. 20 (1969) 230; N. Rott, J. Appl.Math. & Phys. 24 (1973) 54
- [5] C.H. Rode, J.C. Theilacker, Advances in Cryogenic Engineering,Plenum (1988) Vol.33,p.391
- [6] C.H. Rode et al, Advances in Cryogenic Engineering, Plenum (1982) Vol.27, p.769
- [7] H. Burmeister et al, Advances in Cryogenic Engineering, Plenum (1988) Vol.33, p.313
- [8] M. Clausen et al, Advances in Cryogenic Engineering, Plenum (1992)Vol.37A, p.653
- [9] K. Bradzinski et al, Proc. ICEC21 Prague (2006), p.35
- [10] D. Kashy et al, Advances in Cryogenic Engineering, Plenum(1988), Vol.37A, p.577

## 5.10 VACUUM SYSTEMS

### 5.10.1 Requirements for Vacuum Systems

*N.B. Mistry, Y. Li, Cornell U.*

We explore the techniques, problems and solutions in the design of ultra high vacuum (UHV) systems for accelerators. This section concentrates on electron/positron storage rings (including synchrotron light sources (SLS)) where the challenges from synchrotron radiation (SR) induced outgassing are greatest, but includes some treatment of proton rings at cryogenic temperatures.

The basic requirements common to most accelerator UHV systems are: (i) Stored beam lifetimes due to residual gas must be  $> 10$  hours; (ii) good lifetime must be achieved soon after initial startup with stored beam, i.e., a short “conditioning time” is desirable; (iii) the system must be capable of quick recovery after sections are vented for alterations; (iv) “smooth” chamber wall design must minimize beam-induced wakefields; (v) very low pressures must be achieved in and near collider IRs, to minimize detector backgrounds from beam-gas scattering; (vi) the system must be designed with adequate flexibility to allow modification for new development, i.e. new insertion devices, beam ports or other facilities.

At locations in an accelerator vacuum system, where very high electric field (both RF and DC) presents, particulate control during the vacuum components fabrication and installation becomes essential, as even micrometer size particles may become sources of field-emissions.

With good clean assembly procedure and proper choice of materials, thermal outgassing is a minor load on the vacuum system of electron storage rings. SR produces the primary gas load due to desorption from the walls. Local pumping must be provided wherever significant amounts of SR strike the wall or absorbers, as the chamber conductance is usually limited. To avoid higher order mode (HOM) losses, the vacuum chamber must be designed with a smooth continuous profile, with no steps  $> 1$  mm and smooth tapered transitions wherever changes in cross-section are needed. To avoid resistive wall losses the chamber must have good electrical conductance at the frequencies contained in the beam bunch and a thin coating of conductive material (e.g. Cu or Ag) may be necessary. Overheating of sensitive lossy parts of a chamber, e.g. ceramic sections, is

Table 1: Useful formulae &amp; values of physical parameters.

[Units: P(Pa); T (deg.K); M(molecular weight);  $k_B$  (Boltzmann constant) =  $1.38 \times 10^{-22}$  mbar $^{-l}$  $\cdot$ K $^{-1}$ ; the common unit of volume (liter) is used here instead of m $^3$ , and cm $^3$  is used for comparison between R and  $\Phi$ .][<sup>a</sup>] For all gases at 20°C, except for air at 20°C or N<sub>2</sub> at 300K where noted.

| Physical Parameter                              | Equation  | Value <sup>[a]</sup>   |
|---|---|--|
| Number density $n$ (molecules/liter)            | $n = P/k_B T = 7.24 \times 10^{19} P/T$ liter $^{-1}$   | $2.47 \times 10^{17}$ (Pa-liter) $^{-1}$<br>$3.33 \times 10^{19}$ (Torr-l) $^{-1}$ |
| Pressure $P$                                    | $P = nk_B T = 1/3nm \langle v^2 \rangle = 13.8 \times 10^{-18} nT$ Pa   | $4.043 \times 10^{-15} n$ Pa   |
| Average molecular velocity $\langle v \rangle$  | $\langle v \rangle = 4\sqrt{\frac{k_B T}{2\pi M}} = 1.46 \times 10^2 (T/M)^{1/2}$ m s $^{-1}$                           | 465 m s $^{-1}$ (for Air)<br>478 m s $^{-1}$ (for N <sub>2</sub> )                 |
| Mean square mol. velocity $\langle v^2 \rangle$ | $\langle v^2 \rangle = \frac{3k_B T}{M} = 2.49 \times 10^4 T/M$ m $^2$ s $^{-2}$  | $2.2 \times 10^5$ m $^2$ s $^{-2}$ (for Air)                                       |
| Collision rate between molecules (one species)  | $R = \frac{1}{2}n \frac{\langle v \rangle}{\lambda_{\text{mfp}}}; \lambda_{\text{mfp}}(\text{N}_2) \approx (0.61/P)$ cm | $8.6 \times 10^{18} P^2$ s $^{-1}$ cm $^{-3}$<br>(for Air)                         |
| Collision rate on Wall per unit area (flux)     | $\Phi = \frac{1}{4}n \langle v \rangle = (\frac{1}{2\pi M k_B T})^{1/2} P$ s $^{-1}$ cm $^{-2}$                         | $2.87 \times 10^{18} P$ s $^{-1}$ cm $^{-2}$<br>(for Air)                          |

possible due to absorption of HOM energy radiated in another part of the chamber, so enough conducting material and cooling must be provided.

In proton storage rings, high stored currents can drive gas ions into the chamber walls and can lead to positive feedback and pressure runaway unless the ion-induced desorption coefficient is kept below a threshold value (Sec.3.3.6). Large proton rings are likely to have cryogenic vacuum systems, such as in the LHC. These have special problems with appreciable flux of SR that causes desorption of cryogenically adsorbed gases. Special screens may be required to shield cryo-sorbed gases from the SR photons.

### 5.10.2 Units, Conversions and Some Useful Formulae

N.B. Mistry, Y. Li, Cornell U.

There are several books which serve as general references [1] for Vacuum Physics and Technology. These contain detailed expositions of the basics and many useful formulae beside those few given here. A very useful compilation of formulae, tables and nomograms is contained in the vacuum products catalog available from Leybold AG [2]. There are also articles in the literature that survey vacuum system design for accelerators [3].

**Units:** The SI unit of pressure: 1 Pa (Pascal) = 1 Newton m $^{-2}$  (used below unless noted). Other commonly used units: 1 Torr (at 0°C) = 133.3

Pa; 1 millibar = 100 Pa; Atmosphere (std.) = 760 Torr =  $1.0133 \times 10^5$  Pa = 1.0133 bar = 14.696 pounds in $^{-2}$ .

The molecular density  $n$  [molecules/cm $^3$ ] of interest in accelerators, is related to the pressure through the kinetic theory of gases. Formulae in Tab.1 show the various relationships and the values of useful parameters.

Pressures in storage ring systems are in the range of nanotorr ( $10^{-9}$  Torr or  $1.3 \times 10^{-7}$  Pa) with beams present. From Tab.1, the density of air (N<sub>2</sub> equivalent) molecules at a nanotorr is  $3.33 \times 10^{10}$  molecules per liter.

From Tab.1, we note that the gas density equivalent to the desorption of one monolayer of N<sub>2</sub> ( $\sim 10^{15}$  molecule/cm $^2$ ) is  $6 \times 10^{17}$  per liter, the equivalent pressure is about 2.4 Pa ( $1.8 \times 10^{-2}$  Torr)! The time (in seconds) required to form a monolayer on a clean surface of unit stickiness can be estimated as  $T_{\text{mono}} = 3 \times 10^{-4}/P$ .

The wall collision rate per cm $^2$  (flux  $\Phi$ ) given in Tab.1, for a surface with unit stickiness, leads to 11.66 liter s $^{-1}$ cm $^{-2}$  as the value of “pumping speed” per cm $^2$  for air at 20°C, a useful number to remember. In modelling a system, the pump-port area can be assigned a stickiness appropriate to model the effective pumping speed of the pump.

### References

- [1] G.L. Weissler, R.W. Carlson, Ed., Vacuum Physics and Technology, Vol.14 in Methods of

- Experimental Physics, Academic Press (1979); A. Roth, Vacuum Technology, North-Holland (Elsevier Science) (1990); G.F. Weston, Ultrahigh Vacuum Practice, Butterworth (1985)
- [2] A.G. Leybold, Vacuum Technology: Its Foundations, Formulae and Tables, Leybold Vacuum Products Inc., Technical Education Group, 1860 Hartog Drive, San Jose, CA 95131, USA
- [3] N.B. Mistry, AIP Proc. 153 (1987) p.1443; O.Gröbner, CERN Acc. School, CERN-85-19; A. Mathewson, CERN Acc. School, CERN-94-01

### 5.10.3 Conductance and Pressure Profiles

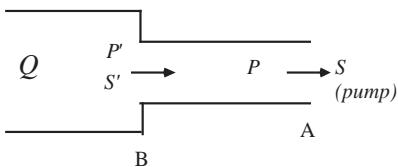
N.B. Mistry, Y. Li, Cornell U.

**Conductance-molecular & viscous flow regimes** At low pressures ( $< 10^{-4}$  Torr) the collision rate between molecules is negligible compared to the wall collision rate (see Tab.1, Sec.5.10.2). The flow is dependent on geometry and density differentials, but not on absolute density. This is the regime of molecular flow.

A rough rule: the molecular-flow regime occurs at the pressure  $P_m$  where the mean-free-path is equal to the diameter of the pipe. For  $N_2$ , at room temperature,  $P_m[\text{Torr}] \sim 5 \times 10^{-3}/d[\text{cm}]$ . Thus for a 5 cm diameter pipe, molecular flow occurs below  $P \sim 1$  mTorr.

At higher pressures, in the viscous flow regime, the flow is determined by the net motion of neighbouring molecules, by the geometry, the pressure gradient, and the absolute pressure (Sec.5.7).

All our discussions below will concern *molecular flow* only. Consider the sketch below:



Throughput: the gas flow rate  $Q$  (in Torr-liter/sec) is related to the mass flow [kg/s] and is constant through a system in equilibrium with no net desorption or adsorption. The pressures  $P$  [Torr] and volume flow rates  $S$  [l/s] are related by

$$Q = PS = P'S' \quad (1)$$

At the output end,  $S$  is the pumping speed, while at intermediate points, the effective pumping speed  $S'$  is determined by the conductance of

the tube in between, defined by Knudsen's Law,

$$Q = C(P'_B - P_A) \quad (2)$$

where  $C$  is the conductance between points A and B, and is a constant, independent of pressure in the molecular flow regime. The units for conductance as for pumping speed are liter/sec, since both are rates of volume flow. Thus the effective pumping speed at point B can be expressed as

$$S' = Q/P'_B = C(P'_B - P_A)/P'_B \quad (3)$$

A rough formula for  $N_2$  in a cylindrical pipe  $L$  cm long, of circular cross-section (radius  $R$  cm):

$$C(N_2) \sim 100R^3/L \text{ [liter/s]} \quad (4)$$

In a multiply connected system, it is useful to consider the analogy to Ohm's Law, and think of  $1/C$  as the resistance,  $Q$  as the current and pressure as the voltage. Thus, conductances in parallel simply add together

$$C = C_1 + C_2 + C_3 + \dots \quad (5)$$

while for conductances in series,

$$1/C = 1/C_1 + 1/C_2 + 1/C_3 + \dots \quad (6)$$

Using Eqs.(1) and (6) we obtain the effective pumping speed  $S'$  at point B,

$$1/S' = 1/S + 1/C \quad (7)$$

Even with infinite pumping speed  $S$ , the pumping at point B is limited by the conductance  $C$ . This is a characteristic feature of most accelerator vacuum systems, where the long narrow beam tube has a low conductance, so that the pumping speed at any intermediate point between pumps is low and "conductance limited". It also explains why large diameter, short tubes should be used to connect high-speed pumps to the vacuum chamber: the conductance must be matched to the speed of the pump.

Simple geometric formulae can be used as approximations to calculate conductances of simple shapes, expressed (in Clausing's method [1]) as *transmission probabilities* multiplying the bare conductance  $C_0$  of the entrance aperture. Table 1 gives conductances of simple tubes and ducts. For finite-length apertures, e.g., small holes in thick walls one uses the Clausing correction [1] factor  $\alpha$  to correct  $C_0$  if accuracy better than 10% is required. For more precise values, or for complex assemblies of irregular shape, it is usual to use Monte Carlo methods to calculate the transmission probabilities [2].

## Sec.5.10: VACUUM SYSTEMS

Table 1: Conductance of Simple Tubes of Uniform Cross section.

[Conductances (units  $\text{m}^3\text{s}^{-1}$ ) are given in terms of  $\langle v \rangle = 478(T/300)^{1/2}(28/M)^{1/2}$  m/s, for  $\text{N}_2$  ( $M = 28$ ) at 300K. Dimensions are in meters:  $R$  (radius);  $L$  (length);  $a, b$  (rectangular dimensions or semi-axes). Conversion:  $1 \text{ m}^3/\text{s} = 10^3 \text{ liter/s.}$ ]

| Tube Shape       | Aperture Conductance, $C_0$          | Tube Conductance   |
|------------------|--------------------------------------|--|
| Cylindrical Tube | $\frac{\langle v \rangle}{4}\pi R^2$ | $[1 + \frac{3}{8}\frac{L}{R}]^{-1}C_0$                     |
| Rectangular Tube | $\frac{\langle v \rangle}{4}ab$      | $[1 + \frac{3}{8}\frac{(a+b)L}{ab}]^{-1}C_0$               |
| Elliptical Tube  | $\frac{\langle v \rangle}{4}\pi ab$  | $[1 + \frac{3}{8}\frac{[(a^2+b^2)/2]^{1/2}L}{ab}]^{-1}C_0$ |

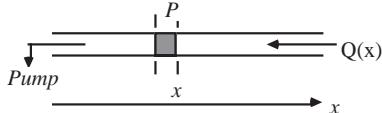
Table 2 Some performance examples

(a) beam lifetimes at normal operating current (low current); (b) change in pressure per change in current at normal operating current; (c) total installed pumping speed of the pumps

| Storage Ring | Energy<br>[GeV] | $I_{\text{tot}}$<br>[mA] | $P_{\text{base}}$<br>[Torr] | $\tau_{\text{i-opr}}(\tau_{\text{low},i})$ (a)<br>[hr] | $S_{\text{tot}}(c)$<br>[l/s] | $dP/dI$ (b)<br>[Torr/mA] |
|--------------|-----------------|--------------------------|-----------------------------|--|------------------------------|--------------------------|
| ALS*         | 1.9             | 400                      | $2.9 \cdot 10^{-10}$        | 6 (8)  | 148,000                      | $2 \cdot 10^{-13}$       |
| CESR         | 5.3             | 340                      | $10^{-9}$                   | 11 (28)  | 95,000                       | $10^{-11} - 10^{-12}$    |
| APS**        | 7               | 100                      | $8 \pm 6 \cdot 10^{-11}$    | 22 (35)  | 552,000                      | $4 \cdot 10^{-11}$       |

\*Courtesy John Thomson, LBNL; \*\* Courtesy John Noonan, John Galayda, ANL

**Pressure in a conductance-limited system** All accelerator and storage ring systems are generally conductance-limited linear systems of mostly uniform cross-section. We define the *specific conductance* of a meter length of chamber:  $c$  liter/m/s. The conductance of length  $L$  is then  $c/L = C$  liter/s.



Here

$$dP(x)/dx = -Q(x)/c \quad (\text{note sign of } Q) \quad (8)$$

$$d^2P(x)/dx^2 = -(1/c)(dQ(x)/dx) \quad (9)$$

For uniform outgassing along the pipe ( $Q(x) = Q_0 + q_0x$ ),

$$P(x) = Ax - \frac{q_0x^2}{2c} + P_0 \quad (10)$$

where  $A$  and  $P_0$  are constants of integration obtained from the boundary conditions.

**Average pressure in a periodic system** A typical problem in a storage ring is to calculate the pressure along a long narrow beam-pipe which outgasses uniformly but is pumped only at discrete intervals along its length. For simplicity, we consider an endless system such as a ring, pumped at regular intervals.

Using Eqs.(8)-(10), we can calculate the average equilibrium pressure and the peak pressure

in a “periodic” vacuum system: a continuous vacuum chamber of specific conductance  $c$  ( $\text{l-m-s}^{-1}$ ), pumped every  $L$  meters by pumps of speed  $S$  [l/s] and base-pressure  $P_0$ . The chamber has a uniform linear outgassing rate  $q_0$  [ $\text{torr-l-s}^{-1}\text{-m}^{-1}$ ] along its entire length,

$$P_{\text{av}} = P_0 + Q_0 \left( \frac{1}{S} + \frac{1}{12C} \right) \quad (11)$$

$$P_{\text{max}} = P_0 + Q_0 \left( \frac{1}{S} + \frac{1}{8C} \right) \quad (12)$$

where  $Q_0 \equiv q_0L$  and  $C \equiv c/L$ .

**Calculating pressure profiles** Due to the complex geometry of accelerator vacuum systems, it is often necessary to model the gas load profile, pumping geometries and gas flow in order to obtain a picture of the pressure profile that can be obtained with a proposed design. This is particularly true for  $e^\pm$  storage rings and colliders incorporating IRs, SR absorbers, crotches and beamlines, together with the sharply varying profile of SR flux and consequent beam gasload. The pressure profile is also crucial in estimating the beam-gas induced backgrounds in the detectors installed at the IP.

Simple 1-D finite-element calculations can be used for a first pass design [3]. Further sophistication is obtained with full 3-D modeling [2] of complex chamber geometries, such as IRs with crossing beamlines incorporating SR masks

and absorbers (gas sources) and discrete and distributed pumping.

In these models, it is important to include the cleaning action of the SR flux in two ways: (i) the decrease in desorption coefficient with accumulating beam dose (see Sec.3.3.6) and (ii) the considerable distributed pumping due to the re-adsorption on an ultra clean chamber wall [4].

The SR gas load for an  $e^\pm$  storage ring of energy  $E_0$  (GeV), beam current  $I$  (mA) and bending radius  $\rho$  (m) can be estimated in terms of the photon flux and the (time-dependent) desorption coefficient  $\eta_{SR}$ . The total photon flux is

$$n_\gamma \left[ \frac{\text{photons}}{\text{m/s}} \right] = 8.08 \times 10^{17} \frac{E_0[\text{GeV}]}{\rho[\text{m}]} I[\text{A}] \quad (13)$$

The estimated gas load is (torr-liter/m<sup>3</sup>)

$$Q = 3 \times 10^{-19} \times n_\gamma \eta_{SR} \quad (14)$$

The desorption coefficient  $\eta_{SR}$  for a typical residual gas (CO) decreases from initial values as high as 0.01 to an ultimate low of about  $2 \times 10^{-6}$  for most chamber materials (aluminum, stainless-steel or copper) after exposure to over  $10^{24}$  photons per meter. For details, see Sec.3.3.6.

The pumping action of ultra-clean chamber walls after exposure to intense fluxes of photons is hard to characterize, since it depends on equilibrium between desorption and re-adsorption and the immediate past history of the surface. However, at CESR, fits to pressure profiles under equilibrium conditions indicate that distributed "wall" pumping ranges from 20 l/s/m to 200 l/s/m in different regions.

Typical thermal outgassing rates can be found in Sec.3.3.6.

**Some  $e^\pm$  storage ring performance** Tab.2 shows vacuum characteristics of three electron storage rings.

## References

- [1] P. Clausing, J. Vac. Sci. Tech. 22 (1972) 111. For a nomogram of conductances using the Clausing factor, see Roth [Ref.1 of Sec.5.10.2], Fig.3.40, p.136
- [2] A. Pace, A. Poncet, Vacuum 41 (1990) 1910; T. Xu, W. Wu, Vacuum 41 (1991) 901; R. Kersevan and J.L. Pons J. Vac. Technol. Sci. A27 (2009) 1017 (MOLFLOW+ program, available from the author)
- [3] Y. Li, CESR note CBN-97-7; J. Howell et al, Calculation of Pressure Distr. in Vac. Systems, ANL/APS Tech. note (unnumbered)
- [4] N. Mistry, Massive Ti Sublimation Pumping in the CESR IR, PAC 97; R. Keservan, Y. Li, N. Mistry, J. Vac. Sci. Tech. A15 (1997) 716

### 5.10.4 Pumping Methods

*N.B. Mistry, Y. Li, Cornell U.*

#### Pumps used on accelerators and storage rings

[1, 2] The vacuum pumps typically used on accelerators and storage ring UHV systems are listed in Table.1, with the ultimate pressure, operating pressure range and pump features. All of these pumps are clean and dry involving no moving mechanical parts needing lubrication, except for some turbo-molecular pumps which use oil bearings instead of magnetic bearings, or those which use oil-lubricated backing pumps.

SIPs, TiSPs and NEG's are the pumps used almost exclusively for routine operation in storage rings. Turbo-molecular pumps are usually used for roughing out large volumes during vacuum system pump-down. However, they should not be connected to the storage ring vacuum system on a regular operational basis, because of the possibility of contamination in case of accidental venting/backstreaming. Discrete cryopumps have not been used extensively in accelerator operations. There is not much experience in operating cryopumps in the presence of intense fluxes of SR where there is the danger of desorbing large quantities of condensed gases. A further problem is the need for isolation of the cryopumps from the beam chamber during periodic warm-ups or power failure. However, in some accelerators (LHC, RHIC, for examples) with large scale use of super-conducting components (such as RF cavities, magnets, etc), beampipe walls at cryogenic temperature provide distributed cryo-pumping. For details of design and operation of pumps of different types, refer to [1, 2].

**Sputter-ion pumps (SIP)** In a SIP, electric and magnetic fields trap electrons, resulting from the Penning discharge, to form a high density electron cloud in the anode cells. Gas molecules are ionized in the electron clouds and the ions are accelerated on to the cathode plates, which are usually made of reactive getter material such as titanium. Fresh cathode (Ti) atoms are sputtered onto the walls and anodes by the ions upon impact on the cathodes. Ions of reactive molecules are chemisorbed on the fresh cathode material, while ions and neutrals of noble gases are physically buried by the sputtered cathode atoms.

Table 1: Pumps used on accelerators and storage rings.

| Pump Type                        | Ultimate Pressure (torr) | Operation Range (torr)  | Major Pump Features  |
|----------------------------------|--------------------------|-------------------------|--|
| Sputter Ion Pump (SIP)           | below $10^{-11}$         | $10^{-4}$ to $10^{-12}$ | <ul style="list-style-type: none"> <li>• Long maintenance-free operating life at low pressures</li> <li>• Poor pumping of rare gases for the normal diode style, but pumps CH<sub>4</sub>; improved pumping of rare gases for noble-diode and triode style pumps.</li> <li>• Suitable for distributed pumping</li> </ul> |
| Titanium Sublimation Pump (TiSP) | below $10^{-11}$         | $10^{-4}$ to $10^{-12}$ | <ul style="list-style-type: none"> <li>• High pumping speed; low cost, easy operation</li> <li>• Requires large surface area for high pumping capacity</li> <li>• Do not pump inert gases such as He,Ar,...CH<sub>4</sub></li> <li>• Suitable for distributed pumping</li> </ul>   |
| Non-Evaporable Getters (NEGs)    | below $10^{-11}$         | $10^{-4}$ to $10^{-12}$ | <ul style="list-style-type: none"> <li>• Compact, easy operation</li> <li>• Limited pumping capacity.</li> <li>• Relatively high cost</li> <li>• Do not pump inert gases such as He,Ar,...CH<sub>4</sub></li> <li>• Suitable for distributed pumping</li> </ul>  |
| Cryogenic Pump (4K)              | below $10^{-11}$         | $10^{-4}$ to $10^{-12}$ | <ul style="list-style-type: none"> <li>• Pumps all gases except He; relatively high H<sub>2</sub> vapor pressure</li> <li>• Large capacity, as compared to getters</li> <li>• Needs auxiliary pump and isolation from main vacuum system for pump re-generation</li> </ul>   |
| Turbo-Molecular Pump             | below $10^{-11}$         | Atm. to $10^{-12}$      | <ul style="list-style-type: none"> <li>• Pumps all gases</li> <li>• Needs backing pump; combination can be started from atmospheric pressure</li> <li>• Oil bearings may contaminate UHV; magnetic bearings preferable; dry backing pump preferable</li> <li>• High cost</li> </ul>                                      |

Table 2: Maximum pumping speed and surface pumping capacity of St101 and St707 NEG strips [8].

| NEG Activation Conditions | Maximum Pumping Speed (liter/sec-cm <sup>2</sup> ) |     |                | Surface Pumping Capacity (torr-liter/cm <sup>2</sup> ) |                    |
|---------------------------|--|-----|----------------|--|--------------------|
|                           | H <sub>2</sub>                                     | CO  | N <sub>2</sub> | CO   | N <sub>2</sub>     |
| St101, 740°C for 45 min.  | 6.7  | 10  | 4.0            | $10^{-4}$  | $6 \times 10^{-5}$ |
| St707, 450°C for 45 min.  | 4.0  | 6.7 | 1.5            | $10^{-4}$  | $6 \times 10^{-5}$ |

In a typical diode pump, both cathode plates are made of titanium. The pumping speed of the normal diode SIP for *noble gases* is usually very small, and the buried noble gases may be subsequently uncovered and released into the vacuum, inducing pressure bursts. Hydrogen molecules, the dominating component in the residual pressure in a typical UHV system, are pumped by diffusion into the bulk of the cathodes. The pumping speed for noble gases may be improved to 25-30% of that for N<sub>2</sub> in the “noble diode”, or differential SIP. In a noble-diode SIP, one of the cathode plates is made of heavy material such as tantalum. The improved pumping of noble gases with the heavy sputtering material is believed to result from the fact that argon or helium atoms are buried deeper due to the higher energy of recoil. The pumping speed for *reactive gases* (CO, N<sub>2</sub>, etc.) in the noble-diode SIP is only 2-3% lower than that in the normal diode pump. Other types of SIP useful for noble gases are the triode SIP [2].

All ion pumps must be shielded from beam-induced RF fields and from scattered SR photons with metallic screens of maximum gas conductance, or performance and particularly the pump ion-current readings will be affected.

**Titanium sublimation pumps** TiSPs are sorption pumps in which Ti is evaporated and deposited on a cold inner wall as the getter surface. Active gas molecules are bound on the Ti film by chemisorption and form stable compounds with the Ti with an unmeasurable low vapor pressure. TiSPs do not pump noble gases and chemically inactive gases such as methane. The pumping speed of a Ti film is proportional to the “sticking coefficient” of gas molecules on the film. The initial sticking coefficient decreases steadily as gas is pumped and active sites become unavailable for impinging molecules, and a new layer of Ti to be sublimated. For all gettable gases, except H<sub>2</sub>, the finite pumping capacity between sublimations depends on not only the geometric surface area, but also the surface roughness of the Ti films. As adsorbed hydrogen readily diffuses into the bulk of the Ti film, the pumping capacity for H<sub>2</sub> is significantly higher, and increases with Ti film thickness.

Typical sticking coefficients at 300K for freshly deposited Ti film are: 0.7 for CO, 0.3 for N<sub>2</sub> and 0.06 for H<sub>2</sub>. Thus, at 300K, the pumping speed of Ti surfaces are 8.2, 3.5 and 2.6 liter/s-

cm<sup>2</sup> for CO, N<sub>2</sub> and H<sub>2</sub>, respectively. More details and references appear in [2], p.200.

Ti sources are usually filaments made of an alloy with a high Ti content, typically 85% Ti and 15% Mo. The filaments are usually heated resistively during the sublimation. For Ti-Mo filaments, it is observed that the Ti sublimation rate may decrease significantly with the number of evaporation cycles at a given evaporation current or power [2, 3]. This should be taken into consideration in long-term TiSP operation.

In storage ring installations, it is very important that a line-of-sight shield is incorporated so that Ti atoms do not enter the beam chamber or contaminate surfaces, e.g. mirrors, ceramic insulators or instrumentation. Massive TiSP pumping has been successfully employed in the IR at CESR [4]. Distributed pumping rates as high as thousands of liters per second per meter have been obtained with freshly sputtered Ti, limited only by the conductance of slots/holes to the beam channel. Careful rf filtering [5] is built into the CESR TiSP control networks, to prevent possible beam-induced rf interference pick-up from Ti filaments being transmitted to sensitive electronics.

**Nonevaporable getters** NEG pumps are also sorption pumps. NEG material [6] is made of special alloys which pump active gases by chemically bonding gas molecules on active atoms or surface sites. The NEGs do not pump noble gases or chemically stable molecules such as CH<sub>4</sub>. As with TiSPs, NEG pumps have a limited pumping capacity.

Activation at high temperature will restore the pumping speed after the NEG surface is passivated by the adsorbed species. The activation temperatures for NEGs are around 400-750°C and NEG material is not evaporated during the NEG re-activation. During re-activation, the adsorbed molecules diffuse into the bulk of the NEG material, leaving empty surface sites for further adsorption of active gases. On the other hand, the adsorption of H<sub>2</sub> is thermally reversible, meaning the adsorbed H<sub>2</sub> will be completely released to the gas phase during NEG re-activation. Besides surface pumping capacity, the ultimate total absorption of a NEG pump is also limited, and the total required lifetime “throughput” must be carefully considered.

SAES Getter (Milan, Italy), has developed and commercialized practical NEGs. Two types of NEGs made by SAES have been used in

storage ring vacuum systems [7] namely St101 (84%Zr-16%Al) and St707 (70%Zr-24.6%V-5.4%Fe). St101 and St707 have very similar pumping performance, in terms of pumping speed and pumping capacity. The activation temperature for St707 (400°C) is significantly lower than that for St101 (740°C), which makes St707 more attractive for accelerator vacuum chambers of smaller size and complicated geometry. The St101 and St707 alloys are usually sintered onto flexible strips of a base metal, such as constantan. However, it has been reported that the low activation temperature feature for St707 may get lost irreversibly whenever heating temperatures higher than 400°C are applied [8]. The activation of the NEGs is done by either resistive heating for the NEG strip, or by conductive or radiative heating for washers of bulk material. Typical maximum pumping speeds and pumping capacities of St101 and St707 strips after activation [8] are listed in Tab.2. Here the pumping capacity is the NEG surface pumping capacity, which is arbitrarily defined as the amount of adsorbed gas to reduce the pumping speed to 10% of the initial pumping speed after activation. Excessive accumulation of hydrogen within either St101 or St707 alloy bulk can lead to embrittlement and particulation due to the change of lattice parameters. It is recommended by SAES not to absorb more than 20 torr-liter/g H<sub>2</sub>.

**SIP-NEG Combination Pumps** SIP-NEG combination pumps, by integrating a compact NEG with a sputtering-ion pump (SIP), has been increasingly used in accelerator vacuum systems. This combination takes advantages of SIP's pumping of non-getterable gases and unlimited pumping capacity, and NEG's high and constant pumping speed (particularly for H<sub>2</sub>) in even deep UHV pressure range. The SIP-NEG combination pumps are commercially available (from Gamma Vacuum Inc., SAES Getters SpA, for example).

**Distributed vs. localized pumping** Pumping has to be provided close to the source of the outgassing, where SR is absorbed on a continuous curving wall. This can be done using linear "distributed pumps". Because of the distributed beam gas load and the restricted conductance along the beam pipes, almost all the storage ring vacuum systems have distributed pumps built into pumping channels or into antechambers which run along parallel to the beam chamber.

In many storage rings, home-made linear sputter ion-pumps have been used in the arcs, using the dipole magnetic field (typically ~2 kG). The structure of these distributed ion pumps (DIPs) [9] is very similar to the SIPs, except that the array is long and narrow. These pumps are installed in the pump-chamber (typically 5 cm by 5 cm in cross-section) extruded parallel to the beam space. The anode cell diameter is tailored to the lower limit of operation of the magnetic field. The typical maximum linear pumping speed is about 100 liter/s-m (N<sub>2</sub>) at pressures between one and ten nanotorr. For increased conductance, the "cells" can be holes in flat anode plates which shape the electric field but allow free flow of gas molecules.

In CESR, intense bunched currents have induced unexpected interactions with the leakage electrostatic fields from the DIPs, causing instability of stored beams (Sec.2.4.13). The scattered photons also induce very large photon current in the DIPs. Both effects are greatly suppressed by lowering the DIP anode voltage to 2 kV [10], without sacrifice of pumping speed. Filamentary metal deposits or "whiskers" arising from long-term operation of DIPs, specially when operating DIPs at lower anode voltage as at CESR, may short the pumps or cause local over-heating of the pump elements. In-situ conditioning of DIPs [9] is usually done by operating pumps at high pressure (> 10<sup>-5</sup> torr), or by dc glow-discharge in N<sub>2</sub>, Ar/O<sub>2</sub> or He/O<sub>2</sub> mixture.

NEG pumps have also been widely used [7] for distributed pumping, where the magnetic fields are very low, making it difficult to operate ion-pumps. Pumping speeds of up to 500 liter/s-m can be achieved using linear strip NEG pumps.

It may be necessary to absorb SR only at discrete specially designed absorbers between magnets due to lack of sufficient distributed pumping. Massive local pumps must be provided right at the absorber with as large a conductance as possible between source and pump. SIPs supplemented with TiSPs or demountable NEG pumps made of accordian strips may also be included. Cryogenic (e.g. liquid helium) pumps may be possible, but careful testing of the concept in the presence of an intense flux of scattered SR photons is necessary. Cryopumps must be isolated by large-aperture gate-valves to allow periodic warm-up.

**'Total' Pumping via NEG Thin Films** It was a revolutionary advance in UHV/XHV technology

by directly depositing non-evaporable getter (NEG) thin films onto the interior walls of vacuum chambers, first developed by Benvenuti *et al* at CERN [11]. After activation of the NEG thin film via heating, the thin film not only greatly reduce thermal and beam induced outgassing, but it also provide so-called 'total' pumping, that is, the entire coated interior wall pumps!

To make the NEG thin films practically useful for accelerator vacuum systems, it is desirable to reduce the activation temperature of the NEG thin films. Ti-Zr-V thin films [11] are most commonly used NEG coating alloy, which can be activated at temperatures as low as  $180^{\circ}\text{C}$  [12]. NEG coatings are used in the LHC and the RHIC vacuum system, in very large scale (hundreds of meters in length). Many SR light sources also employed NEG coating onto insertion device beampipes with very small vertical apertures.

The NEG thin film is usually deposited via standard DC or DC Magnetron sputtering process, using twisted Ti, Zr and V wires as sputtering target. Typical coating thickness is  $1 \sim 3\mu\text{m}$ .

The major drawback of the NEG thin films is their limited pumping capacity. The thin films' pumping performance degrade significantly after more than 10 venting cycles. Before activation of the NEG coated vacuum systems, it is important to bake the system at a temperature below activation temperature (that is, below  $150^{\circ}\text{C}$ ) for sufficient time, so that the NEG films do not prematurely saturate.

## References

- [1] Ref.[1], Sec.5.10.2
- [2] K.M. Welch, Capture Pumping Technology, Pergamon (1991)
- [3] I. Kokelo, Y. Li, PAC 97; CESR note CBN-97-9
- [4] N. Mistry, Massive Ti Sublimation Pumping in the CESR IR, PAC 97; R. Keservan, Y. Li, N. Mistry, J. Vac. Sci. Tech. A15 (1997) 716
- [5] Y. Li, W. Lucas, N. Mistry, CESR note CBN-97-8
- [6] Georgi et al, J. Vac. Sci. Tech. A3 (1985) 417; C. Benvenuti et al, J. Vac. Sci. Tech. A6 (1988) 2528; Georgi et al, J. Vac. Sci. Tech. A7 (1989) 218; C. Benvenuti, F. Francia, J. Vac. Sci. Tech., A8 (1990) 3864
- [7] C. Benvenuti, NIM 205 (1983) 391; LEP Vacuum Group, Vacuum 41 (1990) 1882
- [8] C. Benvenuti, F. Francia, J. Vac. Sci. Tech. A14 (1996) 3278

- [9] Y. Suetsgu, M. Nakagawa, Vacuum 42 (1991) 625; Y. Suetsgu, Vacuum 45 (1994) 133
- [10] Y. Li, R. Keservan, N. Mistry, J. Vac. Sci. Tech. A15 (1997) 2413
- [11] C. Benvenuti, et al, Vacuum 53 (1999) 219-225
- [12] Y. Li and S. Ho, Proceedings of the 2003 Particle Accelerator Conference, p.620; M. Hahn, *et al*, Vacuum 81 (2007) 759-761

## 5.10.5 Instrumentation

*N.B. Mistry, Y. Li, Cornell U.*

**Total pressure measurement** Typical total pressure gauges used in accelerators are listed in Tab.1. Large storage ring vacuum systems are usually divided into sectors, which can be separated by gate valves. Each section should be equipped with at least one high pressure gauge, such as a convection gauge, for monitoring pump-down and for interlocking the vacuum system in case of a very large leak. At UHV pressures, ionization gauges are used. Because of their intrinsic ruggedness, cold cathode gauges (CCGs), specially the inverted magnetron gauges, are increasingly used, in preference to hot cathode ionization gauges. In the presence of background radiation the start-up of CCGs at very low pressure (a common issue) is usually not difficult, due to abundant source of radiation that 'ignite' the gauge. As compared to CCGs, Bayard-Albert gauges (BAG's) have faster response time and better linearity. As a result, BAGs should be used on pressure sensitive components, such as rf cavities, for pressure interlocking. However, BAG's are more sensitive to x-ray interference. For many accelerators, ion currents of SIPs are also used for pressure monitoring. All ion pumps and gauges must be shielded from beam-induced rf fields with copper screens of maximum gas conductance, or performance and particularly the ion-current readings will be affected.

Sensitivity of all ionization based total pressure gauges (including CCG, BAG and Penning

Table 1: Gauge types most used.

| Gauge Type                      | P - Range [Torr]     |
|---------------------------------|----------------------|
| Convection / Pirani             | $1000 - 10^{-3}$     |
| Penning                         | $10^{-2} - 10^{-6}$  |
| Bayard-Alpert ionization        | $10^{-4} - 10^{-12}$ |
| Cold Cathode Magnetron          | $10^{-3} - 10^{-11}$ |
| Cold Cathode Inverted Magnetron | $10^{-3} - 10^{-12}$ |

Gauge) are gas-dependent. Normally these gauges are calibrated to N<sub>2</sub> by the manufacturers. The sensitivity of these gauges to other gases need to be scaled by the ionization probability of the gas relative to that of N<sub>2</sub>. The relative ionization probabilities for common molecules/atoms can be found in most vacuum technology and science books, e.g. [1].

**Residual gas analyzers (RGAs)** RGA is an important part of an accelerator vacuum system. One important function of the installed RGA is to detect small leaks at low pressure with minor interruption to the accelerator operation. An RGA is also essential in analyzing residual gas composition in the accelerator vacuum system in order to understand thermal desorption and SR-induced desorption, as well as to evaluate pumping performance of installed vacuum pumps.

Typical RGAs consist of quadrupole mass filters with mass scanning range at least up to 100 AMU and electron multiplier ion detector having partial pressure sensitivities to below 10<sup>-12</sup> torr.

To interpret an RGA spectrum qualitatively, one must recognize three characteristics: fragmentation patterns, multiply charged ions and isotope ratios. Many organic and inorganic molecules have unique fragmentation patterns (cracking patterns), and these unique fragmentation patterns are used as finger-prints for identifying suspected leaks in the vacuum system, especially those leaks involving interface between the vacuum chamber and sealed cooling or dielectric fluids. Fragmentation patterns for some organic and inorganic molecules can be found in [2] and in online resources, such as [3]. Library spectra provided by most RGA vendors for commonly encountered substances are also very helpful in qualitative spectrum interpretation. Quantitative interpretation of mass spectra, on the other hand, is a rather complicated task, which requires not only knowledge of fragmentation pattern and ionization probability of gases in the vacuum system, but also knowledge about many instrument dependent parameters such as mass filter transmission factor, etc. Fortunately, qualitative RGA spectrum interpretation is often adequate for vacuum system diagnosis.

Typical RGA spectra in a well-conditioned storage ring are usually simple and easy to interpret, at least qualitatively. H<sub>2</sub> usually constitutes more than 80% of the residual gases, while H<sub>2</sub>O, CO and CO<sub>2</sub> are among other species

usually present. As an example, the typical gas compositions in the CESR vacuum system are {Component, (%)}: H<sub>2</sub> (>85), CH<sub>4</sub> (<2), H<sub>2</sub>O (<2), CO (<5), O<sub>2</sub> (<1), CO<sub>2</sub> (<2).

### Practical considerations

Synchrotron radiation effects: Photo-electrons generated by SR can greatly affect the results of ionization gauges and RGAs, so the gauge tubes should always be installed out of the beam plane on the inner radius of the ring, and should avoid line-of-sight to the stored beam, whenever possible. Sensor control electronics should be carefully placed to avoid radiation damage. Lead shields are usually necessary for the on-board electronics. The plastic sheaths of the connecting cable can become hardened and brittle due to radiation.

EMI and RFI: Due to the intense rf fields induced by beam-bunches, it is imperative to shield every ion gauge, ion pump and RGA sensor using copper mesh screens on the port flanges or gaskets, or pressure readout will be wrong.

### References

- [1] Ref.[1], Sec.5.10.2
- [2] Eight Peak Index of Mass Spectra, Vol.1-3, 4th ed. (1991), the Mass Spectrometry Data Centre, The Royal Society of Chemistry; F.W. McLafferty, D.B. Stauffer, The Wiley/NBS Registry of Mass Spectral Data, Vol.1-7, Wiley-Interscience (1989); F.W. McLafferty, D.B. Stauffer, The Important Peak Index of the Registry of Mass Spectral Data, Vol.1-3, Wiley-Interscience (1991)
- [3] URL: <http://webbook.nist.gov/chemistry/nameser.html>

### 5.10.6 Vacuum Chamber Design and Fabrication

*N.B. Mistry, Y. Li, Cornell U.*

#### Chamber geometry and general considerations

Electron rings For e<sup>±</sup> rings, a multi-chambered vacuum vessel is usually required, with spaces for the beam, for pumping, for cooling and heating for bakeout. Various solutions for absorption of SR power and pumping geometries have been used and are described in the literature, with merits and disadvantages for each: (i) SR absorbed on a continuously curved wall with distributed pumping built into the opposite side (the “conventional” solution); simple extrusion of constant shape; difficult to provide enough

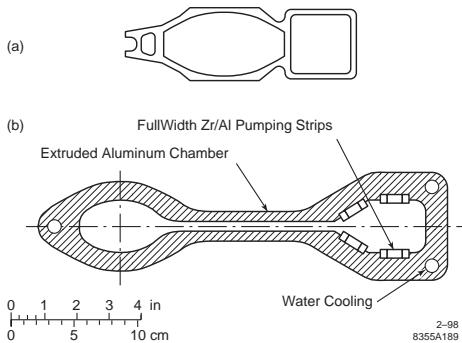


Figure 1: Typical chamber cross sections for electron accelerators.

distributed pumping for intense beams [1]; (ii) SR absorbed on discrete absorbers between bend magnets, with massive discrete pumping; better control of localized heat deposition and massive pumping where most needed; complicated fabrication techniques [2]; (iii) SR absorbed on the wall of an “ante-chamber” separated from the beam chamber by a narrow slot, with distributed pumping on both sides of the beam chamber; the ante-chamber isolates the desorption region from beam to some extent; extrusion difficult, control of distortion required; there arises the question of resonant rf structure effects [3]; (iv) Near the IR, where nanotorr pressures must be maintained in regions of high SR flux, a recent approach [4] uses a variant of (i). The chambers include pumping slots opposite the water-cooled absorber bar, which lead to a pumping plenum with large conductance and surface area. The plenum surface is covered by titanium sublimated by many Ti-cartridges, and provides massive local getter pumping action. Fig.1 illustrates two chamber cross-sections, of type (i) and (iii).

Aluminum and stainless-steel have been commonly used as materials – both are suitable (low outgassing and desorption). Extruded copper-alloy (CuSn) is used in the HERA electron ring [5], mainly to suppress the escape of high energy SR photons outside the chamber. Copper chambers are being used in the B-factories PEP-II and KEK-B. The pros and cons of each material are described below.

**Aluminum Alloys:** Type 6061 and 6063 alloys are commonly used. Easy to extrude complicated, multi-channelled shapes. Good welding characteristics for UHV. High thermal and

electrical conductivity. Very low gas desorption rates, especially with passivated surfaces. Bake-out limited to < 150°C. Multi-chambered shapes have been extruded commercially. Each extruded length is tested for mechanical integrity of the “weld” where the material flows together over bridges in the extrusion die; pumping slots are cut in the wall separating the beam-chamber from the pump-chamber; the chambers are then bent to the correct curvature and the ends machined appropriately. An aluminum chamber may require substantial Pb shielding to absorb scattered high-energy SR photons which produce radiation damage in magnet coils and high levels of ozone and nitric acid in the tunnel air.

**Stainless steel:** Type 304L and 316L stainless steels are commonly used. Simple tubes or complicated flared shapes have been made from machined and welded up parts of SS. Very low outgassing and desorption after treatment. Can be baked to high temperatures. Poor thermal and electrical conductivity, may need conductive coating to reduce resistive wall effects. Caution : Care must be taken to avoid magnetising material in the region of SS welds, where the quality of the magnetic field may be ruined by the chamber permeability.

**Copper and alloys:** Pure copper (OFHC), or CuSn (copper-bronze) [5] if needed for material strength. Good for absorbing very high energy SR avoids escaping radiation. Can be extruded only as single tubes. Multiple chambers must be vacuum brazed or electron-beam welded together. Very good thermal and electrical conductivities. Low outgassing rates possible.

**Proton rings** Proton rings have conventionally used SS or inconel tubes with discrete pumping using flanged pumps. Special degassing techniques are necessary for keeping the ion desorption coefficient below critical threshold for pressure runaway [6] (Sec.3.3.6).

Very high energy (TeV) proton and heavy ion storage rings [7] require further special considerations. A TeV ring with SC magnets may use a cold vacuum chamber operating at LHe temperatures (Fig.2).

In TeV proton rings, a significant flux of SR is generated, leading to photon-induced gas desorption, which is usually not a factor in lower energy proton rings. The chamber wall is an efficient cryopump, but the lightly physisorbed molecules can easily be desorbed by SR photons. Recent designs

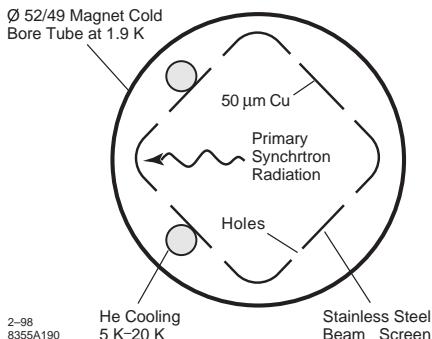


Figure 2: Outline of the LHC vacuum chamber.

[7] call for an inner liner or screen operating at 5K to 10K within a cold bore at 1.9K (superfluid helium). Desorbed gases are pumped on the screen and cold bore and build up condensed layers. As the layers build up, the pressure may become unstable due to ion bombardment, since the desorption yield for a 5 keV  $H_2^+$  ion incident on several monolayer thick condensed hydrogen can be as large as  $5 \times 10^4$  molecules/ion! Further, the vapor pressure of condensed monolayers of hydrogen at 4.2K is about  $10^{-6}$  Torr, which leads to beam-lifetimes of minutes. Thus the beam-tube may need to be warmed up and pumped out before a monolayer can be built up (Sec. 3.3.6). This dictates that the screen and cold bore should be very clean and degassed before installation, to prolong the interval between warm-ups [7]. To minimize image currents, the screen should have high conductivity. However, a low-resistance copper screen cannot be used, as it would be deformed due to large eddy-current forces during a magnet quench (Sec.5.5). A stainless steel screen tube with a thin electrodeposited layer of copper has been proposed. This poses problems of producing a low-desorbing electrodeposited layer.

#### **UHV materials, cleaning, degassing, welding, brazing, assembly**

**UHV compatible materials** [8]: The following list includes common UHV compatible materials.

**Stainless steel:** Austenitic, 300 Series. 304L and 316L preferred for most chamber construction. (L= low carbon, to avoid (chromium) carbide precipitation in heat-affected welding and brazing zones, which can decrease corrosion resistance and mechanical strength.) Type 321 and 347 also used.

**Aluminum:** 6061 and 6063 alloys in T6 condition, suitable for extrusion and welding. Alloys 5454 and 5058 also usable, less susceptible to cracking on bends. Ultra pure aluminum (99.99%) has also been used [9].

**Copper:** Only CD10100 oxygen free, high conductivity grade should be used. The oxygen content should be preferably below 5 ppm. Alloys with 2%-tin (Sn) [5] have been used to provide additional strength at bakeout temperatures.

**Other metals:** Titanium, (pure, aircraft grade), Beryllium-Copper (for spring finger contacts), Noble metals, Kovar, Inconel, Monel, Molybdenum. Dispersion-strengthened copper (e.g., pure copper with a trace of alumina, known as Glidcop®) is useful in absorbers and crotches due to its mechanical strength at high temperature and for its high thermal and electrical conductivity. For example, grade C15715 has a yield str. of 450 MPa at 300°C whereas pure copper has already softened. At 1000°C the C15715 retains a yield strength of almost 300 MPa.

**Transition materials:** Al-SS, Cu-SS and other explosion-bonded transition materials are described below. They require special care in inspection and testing before use.

**Dielectric materials:** Alumina, Beryllia, etc., Sapphire, Quartz, Glass. Some machinable dielectric materials including Macor®, PEEK (Polyether ether ketone) [10].

**Tungsten alloys:** Tungsten alloys can be porous, and should not be used as vacuum walls! These alloys can be used internally in beam-stops and masks, provided they are fully degassed at 1000°C before installation.

**UHV incompatible materials:** Alloys and compounds of Zinc, Cadmium, Lead; and Organic materials, including pump oils.

**Treatment of chambers** The most important need is care in pre-treatment. Cleaning, bake-out and glow-discharge where necessary are crucial to performance. Low-energy injection into GeV  $e^\pm$  rings can prove very difficult because initial outgassing by SR cannot be overcome. Removal of desorbable gases is the prime reason for cleaning of chambers before installation. Initial degreasing to remove machining lubricants and other fluids is usually necessary. It is advisable not to allow any oil-based cutting or machining fluids in manufacture, as these are very hard to remove. Only water based detergent lubricants should be used. An example is CIMCOOL®, manufactured by Cincinnati Millicron.

The most effective treatment after installation in  $e^\pm$  rings is the clean-up (“conditioning”) of the chamber walls by the intense SR beam itself, resulting in low desorption after tens to hundreds of Amp-hours of high-energy (multi-GeV) beam operation [11].

All venting to atmospheric pressure after conditioning should be done using ultra-pure (99.99%) nitrogen. This is essential to obtain quick recovery of UHV [12]. The  $N_2$  serves to passivate the surface, occupying binding sites and denying them to contaminants like  $H_2O$ . For re-assembling or replacing parts of chambers (e.g. rf windows, gauges, etc.) in the tunnel, it is possible to use a portable “clean-room” or shroud around the open port, all the while flowing ultra-pure nitrogen through the chamber, exiting through the open port.

*Component Cleaning:* Most large storage ring UHV systems are made of either SS, Cu or extruded Al. Cleaning usually involves chemical treatments, to remove the thick scaly oxides present from the mill: degreasing of chamber tubes in solvents or vapors; etching away of the surface oxide; washing in distilled water and rapid drying (hot  $N_2$  or ethanol rinse). In most cases, chemical cleaning is done before any welding is done, but after all machining is complete. This avoids entrapment of corrosive chemicals in weld joints and inaccessible corners. The only exception is when full-penetration welds are used exclusively throughout the system. Cleaning methods and formulae are described in [13] and in unpublished reports available from the major laboratories.

Novel methods of cleaning include: Oxygen discharge [14]; dichloropropane treatment [15]; rinse with ozone-rich water [16]; Oxygen discharge is particularly useful in removing hydrocarbons. The dichloro gas treatment is intended to remove adsorbed water while avoiding in situ baking at high temperatures but uses hazardous reagent gases.

*Coating:* Coatings with low secondary electron yield (SEY) are one of effective ways to suppress non-linear beam induced pressure rises beam pipes with positive charged beams (positrons, protons and ions, due to electron cloud buildup). TiN [17], NEG thin films [18] and, more recently amorphous carbon coating [19] are shown to be successful in lower SEY. (Sec.2.4.14).

*Degassing:* Bakeout at high temperature ( $\sim 300^\circ C$  for SS,  $\sim 180^\circ C$  for Al and Cu) is the usual treatment after fabrication, effective in removing adsorbed  $H_2O$  and other lightly bound molecules. Some SS chambers have been made from vacuum fired ( $\sim 1000^\circ C$ ) sheet, to reduce diffused  $H_2$ . Study [20] also shown that ultralow hydrogen outgassing rate ( $\lesssim 10^{-14}$  torr ·liter/sec·cm<sup>2</sup>) may be achieved by vacuum or air bake SS chambers at  $\sim 400^\circ C$  for sufficiently long period. In situations where bakeout in situ is not possible, passing hot  $N_2$  at atmospheric pressure through the sector for several hours may be sufficient to remove adsorbed water, particularly after a controlled exposure to pure nitrogen followed by air.

*Glow Discharge cleaning:* The most effective method for in situ cleaning, since it cleans by actual bombardment of the surface by heavy ions of energy high enough to knock out chemisorbed molecules surviving bakeout. In case of accidental contamination by hydrocarbons, etc. argon/O<sub>2</sub> glow discharge is the only solution [21]. Routine glow-discharge of all chambers is not necessary unless beam-cleaning is inadequate or if the initial outgassing load may overwhelm the installed pumps. Residual argon could be a problem unless completely driven out by bakeout or by pumping with a high ( $10^{-5}$  Torr) pressure of  $N_2$ . The best procedure is to build in an anode for in situ glow-discharge in a completely assembled and baked out system. At CERN a retractable anode wire was installed in the ISR beam chamber. At PETRA and CESR, the anodes of distributed ion pumps in the adjacent pumping channel have been used as discharge anodes.

*Welding and assembly:* Inert gas shielded arc welding (GTAW and GMAW, formerly known as TIG and MIG, processes) and electron-beam welding are commonly used for making all-metal UHV systems. Soldering and gluing are not acceptable! Furnace brazing of small assemblies is possible. An extreme case of vacuum furnace brazing of entire copper chamber lengths was necessary for the HERA electron ring. In most cases, “clean” welding is done after parts have been chemically cleaned, and no chemical cleaning is allowed later since acids and corrosive remnants may get trapped in weld joints. An exception can be made when there are only full penetration welds on all joints, which allows chemical cleaning after welding. “Inside” welds (on the vacuum side) are best, but not always possible.

Appropriate welding techniques and choices of filler material and procedures are best studied from the Welding Handbook, Vol.1 (Technology) and Vols.3 & 4 (Materials and Applications), published by the American Welding Society, Miami, Florida.

Stainless-steel parts are generally easy to weld. Proper weld joint design and preparation is very important. It is crucial to use the correct composition of filler rod for the particular alloy being welded, otherwise the weld will crack upon cooling. Austenitic SS parts that have been previously brazed at high temperature and cooled rapidly can turn to brittle martensite and crack in the weld zone after subsequent welding. Further, the brazing and welding precipitates chromium and nickel as carbides, decreasing the corrosion resistance of the weld-zone. To compensate for this, fillers high in chromium and nickel should be used. [For welding 304L, filler rod type 308L is normally used. To compensate for carbide precipitation after brazing cycles, 309L should be used.]

Welding Al alloys is not as straightforward. The slightest contamination in an Al joint is a recipe for disaster, as inclusions tend to float around, and rewelding is usually a mistake. Hot cracking is prevented by using the proper filler rod for each alloy, e.g. #4043 filler rod for aluminum 6061 or 6063 alloys. Cleaning off the oxide layer is important and “reverse polarity” welding with argon gas helps achieve this during welding. Pulsed or square-wave welding with partial cycles of “forward” and “reverse” polarity may achieve the best balance between penetration and cleaning. The optimum technique for each configuration and alloy should be determined by systematic trial. In short, UHV aluminum welding is an art.

Copper, stainless-steel, niobium and titanium welds (between similar metals) can be made by fusing well-fitting joints with an electron-beam welder, in vacuum. Some electron-beam welds can even be made between dissimilar materials (e.g. titanium and niobium). This technique can be used to weld a thin titanium injection window into a niobium ring. Electron-beam welds are widely used as the main technique in the fabrication of SC rf cavities, where improper welds may become a major cause for limiting the field and the  $Q$ -factor.

Welding dissimilar metals is possible only in rare cases and requires special techniques for each

case. Welding Cu to SS has been quite successful using the GTAW (TIG) process and Cu-Ag eutectic filler wire normally used for vacuum brazing, (BVAg-8: Cu:Ag::28:72). The technique has to be learned by experience, as the eutectic temperature is 780°C, and the filler must be made to wet both base materials! Preheating is a good idea, and so is backup with argon gas inside the chamber.

Transition materials - bonded joints: Transition pieces to SS are needed for attaching standard flanged components to Al chambers. The usual method has been to use Al/SS bonded joints machined out of a large plate or tube of prefabricated bonded material. The Al side is then welded to the Al chamber and the stainless side to a flange or bellows as required. Other clad combinations (e.g. Cu/SS) are also available, useful for Cu chambers.

Prefabricated bonded (clad) material is available in the following forms:

- explosion-bonded Al on SS, with a tantalum or silver interlayer(s); quite expensive, of variable quality unless carefully tested; available in quite large thicknesses (up to 5cm Al/ 5cm SS) [22].
- roll-bonded material, available in thinner clads ( $\sim 1.5\text{cm}/1\text{cm}$ ); less expensive and made in a more controlled way. At CESR we used ordinary aluminum clad stainless-steel frying-pans made by an inexpensive ultra-clean roll bonding process [23].
- friction-welded tubular joints; a reliable method for medium diameters, up to about 2 in. diameter [24].
- DEPI-welding, developed at DESY: direct welding of Al to SS parts with a silver coating; specially trained artisans are needed! (See [1]: PETRA).

The bonded material must be tested thoroughly when received, or one risks a failure of all parts machined from the piece, if the bond is faulty. Mechanical tests, radiography and ultrasonic tests on the plate are used to mark unbonded areas. Heat-cycling tests and leak-tests of *each* machined piece are usually made before the piece is used. The important thing is to test each machined bond-joint (preferably after heat-cycling between 200°C and LN<sub>2</sub> temperature) before welding into a complex chamber. Welding temperatures must be controlled so as not to exceed the maximum tolerable bond temperature ( $\sim 250^\circ\text{C}$ ). Tight fitting copper chill rings around

the bond line should be used while welding, and the bond temperature monitored during welding.

### References

- [1] Various  $e^\pm$  storage rings: CESR: see Ref.[3], Sec.5.10.2; PETRA: C. Falland et al, Proc. 8th. Int. Vac. Congress (1980); PEP: J. Rees, PAC 77, p.1836; KEK Photon Factory: Y. Hori et al, Vacuum 47 (1996) 621; TRISTAN: H. Ishimaru et al, AIP Proc. 236 (1991); LEP: LEP Vac. Group, Vacuum 41 (1990) 1882
- [2] J.M. Lefebvre, Symp. Vac. Sci. & Tech., & SRS Vac. Sys. (1995) Indore, India; M. Renier et al, AIP Proc. 236 (1991) p.71; S.H. Be et al, AIP Proc. 236 (1991) p.103; M. Bernardini, AIP Proc. 236 (1991) p.189
- [3] J.R. Noonan, PAC 97; R.C. Niemann et al, AIP Proc. 236 (1991) p.84
- [4] N. Mistry, PAC 97; R. Keservan, Y. Li, N. Mistry, J. Vac. Sci. Tech. A15 (1997) 716
- [5] R. Ballion et al, Vacuum 41 (1990) 1887
- [6] CERN-ISR-VA/76-27 (1976); A. Mathewson et al, Proc. 7th Int. Vac. Congress (1997) p.1027
- [7] O. Gröbner, Vacuum 47 (1996) 591; also Vacuum 46 (1995) 797; L.P. Hobson, K.M. Welch, J. Vac. Sci. Tech. A11 (1993) 1566
- [8] W. Espe, Materials of High Vacuum Technology, Pergamon (1968); A. Miyahar, Vacuum 41 (1990) 1866
- [9] H. Ishimaru et al, AIP Proc. 236 (1991); H. Iguchi et al. J. Vac. Sci. Tech. A11 (1993) 1708; H. Ishimaru et al, PAC 83, p.2906
- [10] A. Murari, A. Barzon, IEEE Transactions on Dielectrics and Electrical Insulation, 11 (2004) 613
- [11] Ref.[3], Sec.5.10.2
- [12] M. Miki et al, J. Vac. Sci. Tech. A13 (1994) 1760; G. Strasser et al, Vacuum 41 (1990) 1846
- [13] For a review, H.F. Dylla, Vacuum 47 (1996) 647; A.G. Mathewson, CERN Accel. Sch., CERN 94-01; A.G. Mathewson et al, J. Vac. Sci. Tech. A7 (1989) 77; A.G. Mathewson, Proc. Xth Italian Congress on Vacuum Sci. & Tech. (1987) in II Vuoto Vol.XVII, no.3 (1988) 102; J. Herbert, R. Reid, Vacuum 47 (1996) 693; also J. Vac. Sci. Tech. A12 (1994) 1767
- [14] M. Saitoh et al, J. Vac. Sci. Tech. A11 (1993) 2518; J. Vac. Sci. Tech. A12 (1994) 826; J. Vac. Sci. Tech. A13 (1995) 515
- [15] H. Iguchi et al, J. Vac. Sci. Tech. A11 (1993) 1708; K. Tatenuma et al, J. Vac. Sci. Tech. A11 (1993) 1719
- [16] C.K. Chan, et al Journal of Physics: Conference Series 100 (2008) 092025
- [17] M. Sato et al, Vacuum 47 (1996) 753
- [18] B. Henrist, et al Applied Surface Science 172 (2001) 95
- [19] C. Yin Vallgren, et al Proceedings of IPAC10, Kyoto, Japan, p.2033
- [20] C. D. Park, et al, J. Vac. Sci. Technol. A26 (2008) 1166
- [21] H.F. Dylla, JVST A6 (1998) 1276; R.M. Lambert, C.M. Comrie, JVST 11 (1974) 530; A. Mathewson et al, 7th Int. Vac. Congress (1977) p.231
- [22] Some suppliers of explosion bonded material are: SOURIAU PA&E, Wenatchee, WA; Atlas Technologies, Port Townsend, WA
- [23] Some suppliers of roll bounded material are: Clad Products, Boulder, CO, and Klad Inc., Naperville, IL
- [24] Some suppliers of friction welded transitions are: Spinweld Inc., Milwaukee, WI; American Friction Welding Inc., Brookfield WI; Interface Welding, Carson, CA; AIM Inc., Parkton, MD; Chicago Turnrite Co., Chicago IL

### 5.10.7 Special Components in the Vacuum System

*N.B. Mistry, Y. Li, Cornell U.*

**Expansion joints** A “sliding joint” is used for flexible bellows expansion joints between chambers [1]. The sliding joints are necessary to allow thermal expansions of the beam pipes during operations, and to facilitate installation of vacuum chambers. The bare bellows is hidden from the beam by beryllium-copper spring finger contacts which short out the induced rf fields and provide a continuous path for the image charges flowing along the walls. To prevent metal-to-metal “galling” in UHV, the rubbing surfaces are hardened (rhodium plated, at CESR) and silver plated for lubrication. In more recent bellows expansion joints (Fig.1), the “step” is made as small as possible, to reduce the beam-induced RF fields [2].

**Rf-shielded gate-valves** Specially designed gate-valves are used, with rf shielded shuttles that bridge the gap between the flanges when the valve is opened. The RF-shielded gate valves are commercially available from VAT Valves Inc.

#### Beryllium beam pipes for the interaction point

High-energy  $e^\pm$  colliders need a low- $Z$ , small radius, thin beampipe at the IP. This is usually made of Be (extruded seamless or with a brazed seam), less than 1 mm thick, of radius about 2 cm. For high-current machines [3] a double-walled water-cooled Be pipe has been used (Sec.5.6).

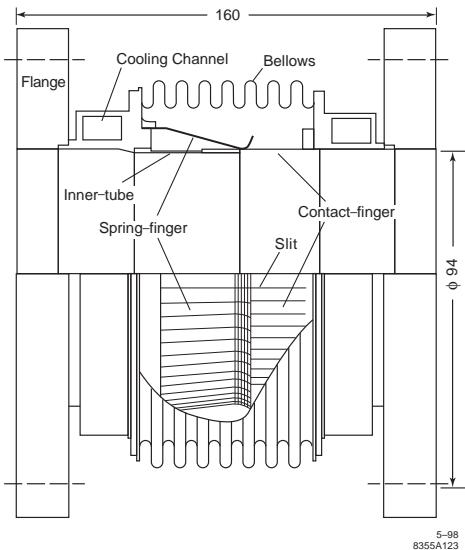


Figure 1: RF shielded sliding joint used at KEK-B.

**Rf cavities, separators and other devices in the vacuum** Rf cavities, electrostatic and rf separators, feed-back electrodes and other such special devices included in UHV systems are described elsewhere in this Handbook (Secs.6.14, 7.2.9). For vacuum design considerations, the choice of low-outgassing materials is paramount. However, discontinuities in shape, e.g. rf cavities, separator plates and tanks, pumping holes and slots, flexible joints, gate valves, all result in induced wake fields and power losses which have to be minimized by proper design. Appropriate cooling must be provided for components that generate or absorb beam-induced rf power. Remember that power radiated at a particular component may be absorbed by (and burn up!) another more dissipative component further down the beam-line.

**Crotches and absorbers** The power density of SR striking the “crotch”, or junction between a beam-line port and the main beam chamber, can be high. Careful design of the vacuum surface and cooling channels is required. A typical crotch design used at CESR [4] for the CHESS x-ray beam lines is shown in Fig.2. The copper absorber is water-cooled and is faced with an electrobonded Be layer which serves to diffuse the intense incident strip of radiation and lower the power density on the copper surface. The cylindrical face was designed to intercept a power density

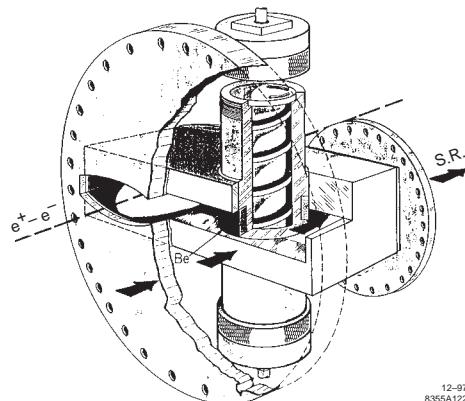


Figure 2: Example crotch assembly.

of  $740 \text{ W/mm}^2$  and was tested with an electron beam welder. Absorbers receiving high power densities can be made of dispersion-strengthened Cu (Glidcop<sup>®</sup>) [5] to withstand the high thermal stresses. However, trials with e-beam welders have shown that even pure Cu can withstand stresses beyond the elastic limit by plastic deformation rather than cracking.

**X-ray beam lines and insertion devices** Detailed descriptions are best obtained from the reports from major Synchrotron Light Sources. Components include crotches, water-cooled collimators, position and profile monitors, fast-acting gate-valves, beam shutters and beam-stops and exit windows [6]. With intense x-ray beams, exit windows must be able to sustain severe thermal stresses due to the absorption of low energy photons in a very thin intense stripe. Direct cooling of the Be window is difficult; some cooling can be obtained with helium gas flowing between two Be windows. Prefilters, e.g. carbon foils, may be necessary to absorb low energy photons in the vacuum line before a Be window and avoid heating the window to dangerous levels.

## References

- [1] Ref.[1], Sec.5.10.6
- [2] Y. Suetsugu et al, Vacuum 47 (1996) 629
- [3] S.D. Henderson, Proc. 8th Meeting Div. Part. Fields of the APS, World Scientific (1994) p.1480
- [4] D. Mills et al, PAC 79, p.3854; S. Yokuchi et al, PAC 93, p.3845
- [5] A. Daoud, et al, Procedia Engineering 2 (2010) 1487

- [6] M.J. Bedzyk et al, RSI 60 (1989) 1460; Q. Shen et al, RSI 60 (1989) 1464; Q. Shen et al, NIM A 347(1994) 609

### 5.10.8 Ceramic Vacuum Chamber Design

*H.L. Phillips, TJNAF*

Herein are summarized the considerations relevant to design of a ceramic vacuum chamber using a particular design as example. Fig.1 shows a schematic elevation. Refs.[1]-[5] provide much useful information on ceramics and brazing technique.

#### Ceramic chamber

##### Alumina ceramic - best choice

- 85% to 97%  $\text{Al}_2\text{O}_3$  - not critical for typical beam line vacuum chamber - in general, higher  $\text{Al}_2\text{O}_3$  content provides greater mechanical strength, increased braze metallization complexity, i.e. higher firing temperature, and somewhat higher cost.
- available
- best braze metallization systems exist for alumina ceramics
- mechanical strength in this range, typically: compression - 200 Kpsi to 300 Kpsi; tension - 20 Kpsi to 40 Kpsi.
- I.D. of ceramic chamber matches I.D. of beam pipe as closely as possible to minimize wakefield generation.
- some fabrication choices: (i) slip cast - solidified slurry coating removed from male mold of chamber inner profile and fired; (ii) isostatic pressing of desired shape and fire - machine pre-fired part from bulk ceramic in green state and fire; (iii) extrusions - inexpensive, watch out for porosity in some products.

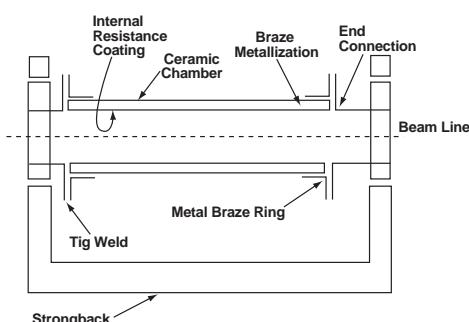


Figure 1: Metallized ceramic vacuum chamber section with strongback.

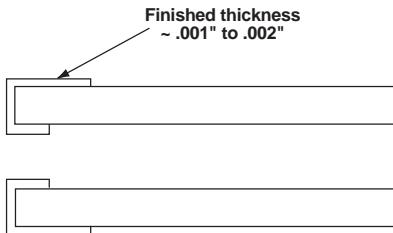


Figure 2: Metalization wrap-around.

The best method of fabrication will depend on length and complexity of the part, the dimensional accuracy required, and how much the ceramic fabricator is willing to do.

- dimensional accuracy - ceramics formed in "green state" larger than desired shape by a "shrink factor" (typically 10-20%). Precision requires wet diamond grinding on fired part.

#### Braze metallization

##### Molybdenum-manganese or tungsten-manganese process - Best Choice

- Mo-Mn suspension - brushed, sprayed or screened on ceramic → fired → nickel plate on fired metallizing → sintered in  $\text{H}_2$  yielding strong, brazeable surface

- interface strength in tension approaches or exceeds tensile strength of ceramic

- metallization should wrap around end of ceramic to contact resistive coating added later (see Secs.6.15.1, 6.15.2)

#### Metal braze rings

- Function - Provides low stress hermetic seal to ceramic - must be thin (typically 0.020" to 0.050") - Provide flexible transition between ceramic and massive metal parts

- Must match expansion coefficient of ceramic reasonably well - "Kovar" is a good choice - welds well to stainless steel.

- Design clearance to metallized ceramic - typically 0.001" to 0.002"

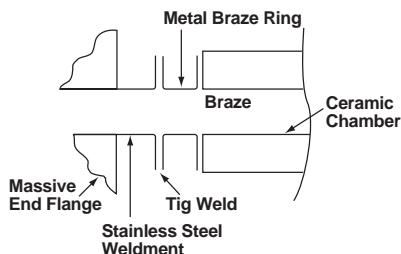
- Braze details: - use vacuum tube grade alloys - typical choices: silver-copper eutectic, gold copper, palladium-silver-copper - Kovar must be nickel plated if silver bearing alloys are used - sulfamate or Watts type nickel will work - electroless may not.

- For large seals, kovar may need to be restrained with many wraps of molybdenum or tantalum wire during braze cycle to control gap.

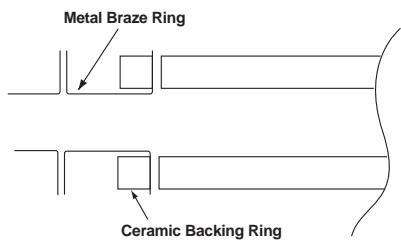
## Sec.5.11: ALIGNMENT

End Flange Options - providing smooth interior and low wakefield generation

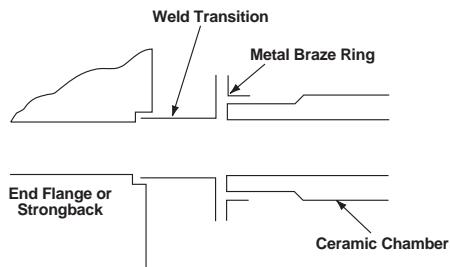
*Example 1 Option no.1:*



*Example 2 A better system - shear stress halved using ceramic backing ring:*



*Example 3 The best system is:*



Advantages: -seal is in compression - low-est failure risk; Disadvantages: - O.D. of ceramic must be locally ground to shape.

Metal braze rings of either circular or non-circular shape may be easily deep-drawn using inexpensive aluminum dies. Brazes to the flat end face of the ceramic tube, as in examples 1 and 2, are conveniently done with braze foils between the ceramic and metal ring. A braze to the O.D. of the ceramic, as in example 1, is conveniently done with braze wire placed at the corner of the braze ring in contact with the ceramic metallization wicking into the joint when liquid. The

volume of braze wire used is sufficient to fill the crack and provide a fillet on either end of the joint. To control the volume of the crack during brazing which increases due to differential thermal expansion many turns of wire of Mo, Ta, or W is often wrapped tightly around the kovar ring.

### Strongback

- Required to eliminate the transfer of torques, moments and longitudinal forces from the beam pipe to the ceramic chamber.
- Differential expansion between ceramic and strongback can be somewhat accommodated by the flexure of the thin radial weld flanges.
- For more severe expansion, a sliding joint must be included in series with the ceramic chamber within the strongback.

## References

- [1] W.H. Kohl, Handbook of Materials and Techniques for Vacuum Devices, AIP (1994)
- [2] F. Rosebury, Handbook of Electron Tube and Vacuum Techniques, AIP (1993)
- [3] Engineered Materials Handbook, Vol.4, Ceramics and Glasses, ASM (1991)
- [4] M. Schwartz ed., Brazing, ASM (1990)
- [5] D. Richerson, Modern Ceramic Engineering, Marcel Dekker Inc.

## 5.11 ALIGNMENT

*R. Ruland, SLAC*

**General** The challenge facing the alignment team is to translate a theoretical accelerator layout designed in Cartesian space into a physical ring in geocentric space - to transform a list of theoretical coordinates into a physical system in which each component lies at its design location to within a specified tight tolerance. The task of positioning magnets can be broken down into six major sub-tasks:

(i) Survey Reference Frame: A survey coordinate system appropriate to the project site and size needs to be defined. Monuments are established to represent this reference grid.

(ii) Layout Description Reference Frame: The beam line is designed and specified in a lattice coordinate system. Coordinate transformations, including rotations and translations, need to be defined to relate this to the survey reference frame.

(iii) Fiducialization: The fiducialization of components relates their effective magnetic centerline to reference points that are accessible to subsequent survey measurements.

(iv) Pre-alignment of Girders (optional): After components and vacuum chambers are mounted on a girder, they are aligned relative to a girder coordinate system.

(v) Absolute Positioning: Components are aligned with respect to the global reference grid.

(vi) Relative Positioning: Local tolerances are achieved by the relative alignment of adjacent components.

**Survey reference frame** The goal is to define a mathematical model of the space in which the surveyor takes his measurements and performs his data analysis.

Surveying space Ancient civilizations realized that the earth is round [1]. It is actually of a more complex shape, the modeling of which is not easy. The geoid is the reference surface described by gravity; it is the equipotential surface at mean sea level that is everywhere normal to the gravity vector. The geoid is non-symmetric and its mathematical description non-parametric, rendering it unsuitable as a reference surface for calculations. The ellipsoid is the regular figure that most closely approximates the shape of the earth. Being a regular mathematical figure, it is the surface on which calculations can be made. Account must be taken of the discrepancy between the ellipsoid and the geoid. The deflection of the vertical is the angle of divergence between the gravity vector (normal to the geoid) and the ellipsoid normal. For a survey-engineering project, it is sufficient to define a best-fit local ellipsoid that minimizes discrepancies only in the local area. All survey measurements made in reference to gravity must be reduced to the ellipsoid before computations can proceed. The equations can be found in most of the geodetic literature [2]. Computations with ellipsoidal coordinates are complex. Therefore, it is desirable to project the spheroidal coordinates into a local Cartesian coordinate system. However, it is not possible to project a spherical surface on to a plane without creating distortions. But since these distortions can be mathematically modeled, it is possible to correct derived relationships, such as distances, angles, or elevations. It should be pointed out that for smaller projects a Cartesian coordinate system will suffice as long

as the vertical coordinate is corrected for the curvature of the earth.

Survey networks In order to achieve the absolute position tolerance and any circumference requirements, a surface network with pillar-type monuments is usually established. The coordinates of these monuments are in most cases determined using Global Positioning System (GPS) methods [3, 4]. Position accuracies of a few mm for small projects to few cm for larger size projects can be expected. The tunnel net extends the coordinate system reference into the immediate vicinity of the components [5]. It is usually tied to the surface network by optically or mechanically centering a tripod-mounted translation stage on the surface over a monument in the tunnel through a survey shaft. Modern instrumentation yields plumbing accuracies of 1:100,000. Tunnel nets are commonly established as longitudinal networks. While in the past forced centered instruments in trilateration/triangulation mode were used to determine the monument coordinates, today's total stations, laser trackers [6], or the total station like tracker hybrid Leica AT401 are best operated in free-stationing mode. This approach greatly eases the complexity of the monument and forced-centering-system designs resulting in significant cost savings and improved accuracies. Relative monument accuracies of 100 m and below can be reached; the obtainable absolute accuracy is directly correlated to the surface net position accuracy.

**Layout description reference frame** The layout description of every machine component is given in the design lattice. Position tolerances are determined for each magnetic component and are attached to the lattice specifications. The individually specified parameters are usually the maximum permissible displacements in the direction of the three coordinates and the associated rotations. The tolerance specifications should distinguish between absolute and relative positioning [7]. Surveying measurements will show a typical Gaussian distribution, including entries outside the chosen confidence level. Achieving the equivalent of the mathematical truncation requires a means to identify "outliers" and a method to add independent redundant observations. Traditionally, the stochastic computations in surveying are based on a  $1\sigma$  confidence level. The relationship between the surveying and lattice coordinate systems is defined as a transformation matrix [8].

**Fiducialization** The fiducialization process relates the effective internal electromagnetic axis of a component to external marks that can be seen or touched by instruments. It is these reference marks that are then aligned onto their nominal coordinates. Magnets are mostly made with ferromagnetic poles, which are traditionally used as the references for external alignment fiducials [9]. A better way to establish fiducials is to use magnetic field measurement [10]. A summary of the most common opto-mechanical and electro-mechanical techniques can be found in [11]. Typically, opto-mechanical methods can yield fiducialization accuracies of 25-75  $\mu\text{m}$ , electro-mechanical fiducialization accuracies of 5-25  $\mu\text{m}$  have been reported [12].

**Pre-alignment of girders** Girders are commonly used to support components and the vacuum chamber of one lattice cell. These girders are pre-assembled and pre-aligned in a factory before they are transported into the tunnel.

**Absolute positioning** The absolute positioning activities can be subdivided into four steps:

Step 1: Blue line survey on the tunnel floor

In preparation for the installation of the support systems, a “blue line” survey is performed to lay out the anchor bolt positions. This is most efficiently done taking polar measurements with a laser tracker or a total station like the Leica TDRA6000 [13]. Without too much effort, a few mm type lay-out precision with respect to local traverse points can be achieved.

Step 2: Rough absolute positioning of components in tunnel After the blue line survey, the anchors are set and the pre-aligned girders or components installed, but with their anchor bolt nuts only “hand tight”. The components’ adjustment systems are set to midrange to ensure that the full adjustment range remains available for fine positioning. The components are rough aligned by tapping them into position utilizing the slack between anchor bolts and support structure. Actual coordinates are determined and compared to ideal coordinates, yielding the position adjustment values.

Step 3: Fine absolute positioning of components The horizontal and vertical positions are set relative to the tunnel monument system; this is usually carried out using the real-time feed-back of a laser tracker. In contrast to Step 2, this time the mechanical adjustment systems are used to move the components accurately. The achievable

absolute component position accuracy is directly correlated to the absolute position accuracy of the reference tunnel network points. For small projects, one can expect to do better than 5 mm, for large size projects the uncertainty might be in the cm range. However, the relative position uncertainties over a short span could still be in the 100-500  $\mu\text{m}$  range.

Step 4: Quality control survey After the absolute positioning of girders and components is completed, a re-survey is conducted to verify the results.

**Relative positioning (smoothing) for large accelerators** The accuracy obtained in the absolute positioning step is the quadratic sum of many random errors plus the linear sum of any residual systematic errors. The typical error envelope for the absolute alignment of a beam line is cigar-shaped; it is at a minimum (but never zero) at the control points and grows to reach a maximum midway between two successive control points. The measured reference line oscillates somewhere within this error envelope. Its absolute position cannot be pinned down any more precisely than the size of the error envelope, and deviations within this envelope are statistically insignificant. However, within this absolute error envelope, relative errors between adjacent magnets should be smaller: the major error sources affect equally the positioning of adjacent components with the result that relative alignment accuracies are significantly higher than absolute alignment accuracies. Consequently, successive surveys will reveal reference lines of different shapes whose absolute position floats around randomly within the cigar-shaped error envelope. Because of this, the absolute positioning technique is not well suited to achieving a final position tolerance. To overcome this problem, techniques were developed to separate relative displacements from the absolute trend curve — techniques referred to as “smoothing” [14]. Smoothing accuracies of less than 100  $\mu\text{m}$  have been reported. After smoothing, the distribution of residuals needs to be examined by Fourier decomposition type analyses to ensure that no significant amplitudes occur at the betatron frequency. It should be pointed out that smoothing techniques cannot be applied to machines where some components need to be placed into a specific relationship to other components, e.g. insertion device with respect to photon beam line.

**Survey and alignment toolbox** The typical contents of an accelerator surveyor's toolbox is listed in [15]. Many surveying textbooks and technical papers cover these instruments and software tools [16].

### References

- [1] W. Torge, Geodesy, deGruyter (2001)
- [2] A. Leick, GPS Satellite Surveying, Wiley (2004), p.188, or Jackson, op. cit., p.84
- [3] R. Ruland, A. Leick, Proc. 1st Symp. on Precision Positioning with GPS (1985) p.483
- [4] A. Leick, op.cit.; B. Hofmann-Wellenhof et al., GPS - Theory and practice, Springer (2001) 5th edition; B. Hofmann-Wellenhof et al., GNSS Global Navigation Satellite Systems: GPS, GLOASS, Galileo, and more, Springer (2008)
- [5] F. Löffler, W. Schwarz, Proc. 1st Int. Workshop on Acc. Alignment, SLAC-375 (1989) Fig.25, p.127
- [6] Laser trackers are manufactured by API Automated Precision Inc., FARO Technologies Inc., and Leica Geosystems AG
- [7] R. Ruland, Proc. 5th ICFA Beam Dynamics Workshop, SLAC-PUB-5672 (1991)
- [8] W. Oren, R. Ruland, Proc. 45th ASP-ACSM Convention, SLAC-PUB-3542 (1985) p.338
- [9] H. Friedsam et al, Proc. 1st Int. Workshop on Acc. Alignment, SLAC-375 (1989) p.206
- [10] C.M. Fortgang et al, Proc. Linac Conf. (1990) p.48; Y. Levashov, Z. Wolf, SLACLCLS-TN-05-10 (2005); F. Löffler, Proc. 1st Int. Workshop on Acc. Alignment, SLAC-375 (1989) p.232; G.E. Fischer et al, ibid., p.213; H. Brück et al., Kern-tech 56 (1991) p.248ff; Z. Wolf et al., SLACLCLS-TN-07-7 (2007); A. Jain et al., 10th Int. Workshop on Acc. Alignment, KEK (2008)
- [11] R. Ruland, CERN Acc. School, SLAC-Pub 7447 (1997)
- [12] Z. Wolf et al., op.cit.; A.Jain, 16th Int. Mag. Meas. Workshop, Bad Zursach (2009)
- [13] R. Ruland, Proc. 3rd Int. Workshop on Acc. Alignment (1993) p.101
- [14] R. Ruland, Accelerator and Transport Line Survey and Alignment, ibid.; M. Mayoud, Proc. 1st Int. Workshop on Acc. Alignment (1989), SLAC-375, p.138; H. Friedsam, W. Oren, SLAC-375, p.152; W. Schwarz, Proc. 2nd Workshop on Acc. Alignment (1990) p.267;
- [15] R. Ruland, in Synchrotron Radiation Sources – A Primer, Ch.11, World Scientific (1994) p.300
- [16] F. Deumlich/R. Steiger, Instrumentenkunde der Vermessungstechnik, Wichmann (2002); F. Löffler et al., Handbuch Ingenieurvermes-

sung - Maschinen- und Anlagenbau, Wichmann (2002); P. Richardus, Project Surveying, A.A. Balkema (1984); H. Kahmen, W. Feig, Surveying, de Gruyter (1988); H. Kahmen, Vermessungskunde, deGruyter (2006); P. Kissam, Optical Tooling for Precise Manufacture and Alignment, McGraw-Hill (1962); B. Witte, H. Schmidt, Vermessungskunde und Grundlagen der Statistik für das Bauwesen, Wittwer Verlag (2006); W. Schwarz, ed., Vermessungsverfahren im Maschinen- und Anlagenbau, Schriftenreihe des DVW 13, Wittwer (1995); A. Grün/H. Kahmen, eds., Proc. of Symposia: Optical 3-D Measurement Techniques, Wichmann, 1989, 1993, 1995, 1997, 2001, 2003, 2005, 2007, 2009; Proc. of the International Workshops on Accelerator Alignment; <http://www-conf.slac.stanford.edu/iwaa/pastConf.html>

### 5.12 MAGNET SUPPORTS AND ALIGNMENT

*G. Bowden, SLAC*

A review of magnet support design can be found in [1].

**Support geometry** Supports fix accelerator components in 3-D space along the beam line. Magnets typically range in size from 0.1 m to 10 m, weigh 10 kg to 10 tons and are assumed rigid objects. They have 6 degrees of freedom (DOF):  $X$ ,  $Y$ ,  $Z$ , pitch, roll, yaw. Stability demands that supports fix all 6 DOFs but to avoid distortion such as twisting and bending, supports must not constrain more than these 6 DOFs. Such ‘kinematic’ supports with six constraints matching six DOFs take 2 forms for accelerator beamline components as shown in Figs.1 and 2.

Both the six strut system and the cone-groove-pad supports are kinematic supports making contact with the magnet at 6 points. The six strut system uses struts with commercial spherical ball joints at each end to insure that they carry only axial loads. Strut length adjustment is by differential screw or a shim cartridge built into the strut. Adjustment of all 6 DOFs can be made independent and orthogonal by arranging support struts at right angles to each other. Joints should have a slight interference fit to avoid free-play, especially in those struts which do not carry weight. The 6 strut kinematic mount can be adapted to almost any component configuration.

Cone-groove-pad kinematic mounts use gravity to fix the magnet in a unique position

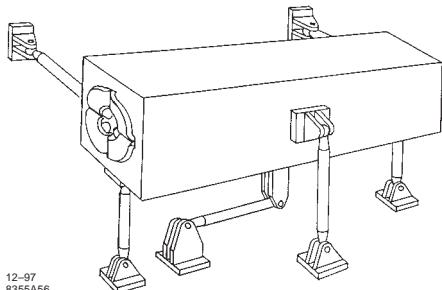


Figure 1: Six strut mount.

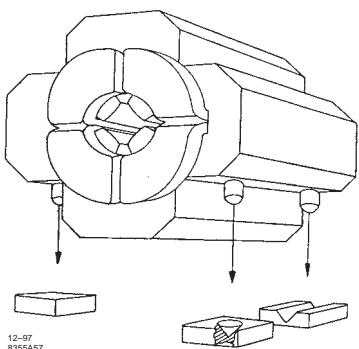


Figure 2: Cone, groove and pad mount.

determined by cone, groove and pad supports which are independently adjustable in  $X, Y, Z$ . For stability and reduced thermal displacement, the cone-groove-pad plane should be near the beam line. This mount geometry is compact and intrinsically free of play. Variations of this mount geometry have been used to construct remote controlled cam driven kinematic positioners for beamline magnets. Five roller bearing cams support the magnet or accelerator section. The 6th DOF; position along the beam line can be fixed by a flexible link. Cam lift is typically 1 mm and positioning to  $< 1\mu\text{m}$  can be obtained in 5 DOFs by rotation of internal cam shafts. Positioning is intrinsically kinematic and bounded. The relations between cam shaft angles and magnet position are transcendental functions but easily evaluated by accelerator control system computers. Fig.3 shows a 5-DOF positioning mount. Fig.4 shows a 3-DOF mount using common shafts for 2 pairs of cams. A detailed description of remote cam positioning mounts can be found in [2].

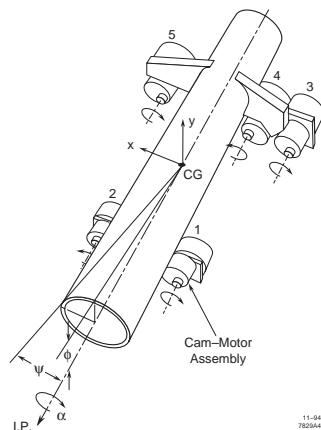


Figure 3: Five-DOF support.

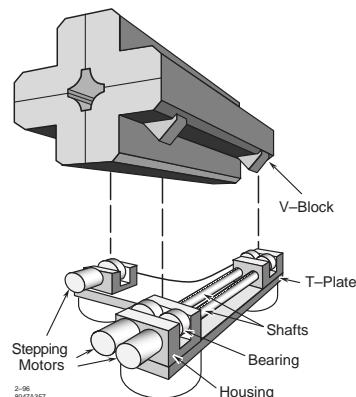


Figure 4: Three-DOF support.

**Support vibration** (See Sec. 5.14) Support should be sufficiently stiff that natural resonant frequency is above 15 Hz. Soft supports with low resonant frequency are susceptible to microseismic vibration and resonate with local mechanical equipment. An approximate natural frequency can be estimated from structure deflection under gravity:

$$f_0 = \frac{1}{2\pi} \sqrt{\frac{g}{\delta}} \quad \text{Hz} \quad (1)$$

Here  $\delta$  (meters) is static deflection of mode under force of gravity  $g = 9.8 \text{ m/s}^2$ .

**Thermal distortion** Thermal misalignment of components is a function of both thermal expansion  $\gamma (\Delta l/l/\text{C})$  and transverse thermal gradients which depend on material thermal conductivity  $\kappa$ . Materials with a low ratio  $\gamma/\kappa$  will

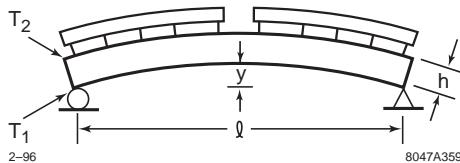


Figure 5: Effect of thermal distortion.  $y = \frac{1}{8} \frac{\Delta T}{\alpha c} l^2 [T_2 - T_1]$  m.

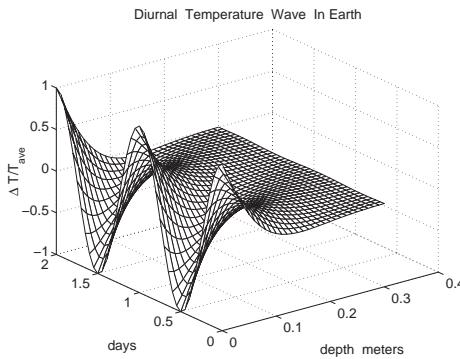


Figure 6: Thermal diffusion in earth.

have low thermal distortion. The simple case of a beam bending under transverse thermal gradient is shown in Fig.5.

**Thermal diffusion** Diurnal heating/cooling cycles diffuse into support structures and accelerator housings. Cyclic temperature excursions decrease and lag in phase with increasing depth of material is shown for earth in Fig.6. One dimensional diffusion of sinusoidal heating with period  $\tau_0$  (days) for material with thermal diffusivity  $\alpha = k/\rho c$   $\text{m}^2/\text{day}$  is

$$\frac{\Delta T(x, t)}{T_{ave}} = \cos \left( \sqrt{\frac{\pi}{\alpha \tau_0}} x - 2\pi \frac{t}{\tau_0} \right) e^{-\sqrt{\frac{\pi}{\alpha \tau_0}} x} \quad (2)$$

Some representative material diffusivities are

| material  | $\alpha$ ( $\text{m}^2/\text{day}$ ) |
|-----------|--------------------------------------|
| earth     | .012                                 |
| concrete  | .044                                 |
| sandstone | .089                                 |
| iron      | 1.75                                 |
| aluminum  | 6.16                                 |

## References

- [1] R. Ruland, in *Synchrotron Sources – a Primer*, World Scientific, Ch.11 (1994) p.234
- [2] G. Bowden et al, NIM 368A (1996) 579

## 5.13 GROUND VIBRATION

C. Montag, BNL, J. Rossbach,  
University Hamburg

### 5.13.1 Basics

The variance  $\sigma^2$  of a ground motion  $z(t)$  ( $z$  represents transverse coordinate, either  $x$  or  $y$ ) can be calculated either from its time signal or from its power spectrum  $p(\omega)$ ,

$$\sigma^2 = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} z^2(t) dt = \int_{-\infty}^{\infty} p(\omega) \frac{d\omega}{2\pi} \quad (1)$$

with the power spectrum

$$p(\omega) = \lim_{T \rightarrow \infty} \frac{1}{T} \left| \int_{-\frac{T}{2}}^{\frac{T}{2}} z(t) \exp(-i\omega t) dt \right|^2. \quad (2)$$

The dimension of  $p(\omega)$  is length<sup>2</sup>/frequency.

The mutual power spectrum of two signals  $z_i(t)$  and  $z_j(t)$  is defined as

$$p_{ij}(\omega) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} dt \int_{-\frac{T}{2}}^{\frac{T}{2}} dt' F(i, j, \omega, t', t) \quad (3)$$

with

$$F(i, j, \omega, t', t) = z_i(t') z_j(t) \exp[i\omega(t' - t)].$$

The correlation (complex quantity)

$$\gamma(\omega) = \frac{\langle p_{ij}(\omega) \rangle}{\sqrt{\langle p_{ii}(\omega) \rangle \langle p_{jj}(\omega) \rangle}} \quad (4)$$

contains information on the relative phase of the motion as a function of frequency.  $|\gamma(\omega)|$  is called the coherence.

For the power spectrum of relative motion of two points  $s_1$  and  $s_2$ , separated by  $\Delta L$  we get

$$\begin{aligned} \rho(\omega, \Delta L) &= \\ \lim_{T \rightarrow \infty} \frac{1}{T} &\left| \int_{-\frac{T}{2}}^{\frac{T}{2}} (z(t, s_1) - z(t, s_2)) \exp(-i\omega t) dt \right|^2 \\ &= p_{11}(\omega) + p_{22}(\omega) - p_{12}(\omega) - p_{21}(\omega) \end{aligned} \quad (5)$$

with variance

$$\begin{aligned}\sigma^2(\Delta L) &= \int_{-\infty}^{\infty} \rho(\omega, \Delta L) \frac{d\omega}{2\pi} \\ &= \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} [z_1(t) - z_2(t)]^2 dt.\end{aligned}\quad (6)$$

For the extraction of some model parameters from measured data, a 2-D power spectrum of  $z(t, s)$  can be introduced [1],

$$P(\omega, k) = \lim_{T \rightarrow \infty} \lim_{L \rightarrow \infty} \frac{\left| \int_{-\frac{L}{2}}^{\frac{L}{2}} \int_{-\frac{T}{2}}^{\frac{T}{2}} z(t, s) \exp(-i\omega t - iks) dt ds \right|^2}{TL} \quad (7)$$

with

$$p(\omega) = \int_{-\infty}^{\infty} P(\omega, k) \frac{dk}{2\pi}$$

$$\rho(\omega, \Delta L) = \int_{-\infty}^{\infty} P(\omega, k) \cdot 2[1 - \cos(k\Delta L)] \frac{dk}{2\pi}. \quad (8)$$

Therefore,

$$\sigma^2 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P(\omega, k) \frac{d\omega}{2\pi} \frac{dk}{2\pi} \quad (9)$$

$$\begin{aligned}\sigma^2(\Delta T, \Delta L) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} \frac{dk}{2\pi} P(\omega, k) \\ &\cdot 2[1 - \cos(\omega\Delta T)] \cdot 2[1 - \cos(k\Delta L)].\end{aligned}\quad (10)$$

A useful model of the uncorrelated part of ground motion is given by the *ATL* rule [2]:

The variance  $\sigma^2$  of the uncorrelated motion of two points separated by  $\Delta L$ , observed after a time interval  $\Delta T$  (assuming  $z_1 - z_2 = 0$  at  $t = 0$ ) is proportional to the distance  $\Delta L$  and the time interval  $\Delta T$ ,

$$\sigma^2(\Delta T, \Delta L) = A \cdot \Delta T \cdot \Delta L \quad (11)$$

with corresponding power spectrum

$$\rho(\omega, \Delta L) = \frac{A \cdot \Delta L}{\omega^2}. \quad (12)$$

*ATL*-like motion has been confirmed in numerous measurements of ground motion [3] and beam orbits in accelerators [4].

### 5.13.2 Measurements

Since measured data consist of real numbers only, there is no additional information from negative frequencies:  $p(\omega) = p(-\omega)$ . Thus a slightly different definition of the power spectrum is useful:

$$p(\omega) = \lim_{T \rightarrow \infty} \frac{2}{T} \left| \int_{-\frac{T}{2}}^{\frac{T}{2}} z(t) \exp(-i\omega t) dt \right|^2 \quad (1)$$

with

$$\sigma^2 = \int_0^{\infty} p(\omega) \frac{d\omega}{2\pi}, \quad (2)$$

i.e.  $p(\omega)$  is defined for positive frequencies only.

The rms-value in the frequency band  $f_0$  to  $f_1$  can be calculated by integrating the power spectrum over this frequency band:

$$\sigma(f_0 < f < f_1) = \sqrt{\int_{2\pi f_0}^{2\pi f_1} p(\omega) \frac{d\omega}{2\pi}}. \quad (3)$$

Using the coherence  $|\gamma(\omega)|$ , the ratio  $\beta(\omega)$  of correlated to uncorrelated ground motion can be computed as

$$\beta(\omega) = \frac{|\gamma(\omega)|^2}{1 - |\gamma(\omega)|^2}. \quad (4)$$

For a compilation of vertical ground motion spectra at various accelerator sites, see Fig.1 [5].

Characteristics of all spectra:

- Microseismic peak around 1/7 Hz, caused by ocean waves hitting coasts (the “7-second hum”).
- Broad peak around 2.5 Hz. Its origin is explained by either some bandpass characteristics of the ground surface or as a result of cultural noise and damping both increasing with frequency, but with damping becoming effective at slightly higher frequencies.
- cultural noise beyond  $\approx 5$  Hz, caused by traffic, heavy machinery, etc.

### 5.13.3 Instruments

**Piezoelectric accelerometers** A piezo crystal is mounted to the sensor housing from one side; the other side is loaded with a seismic mass. Therefore, the crystal acts as a kind of spring which is sheared when the sensor is accelerated in its sensitive direction. Below the mechanical resonance frequency (usually several kHz), the output charge  $C$  is proportional to the acceleration

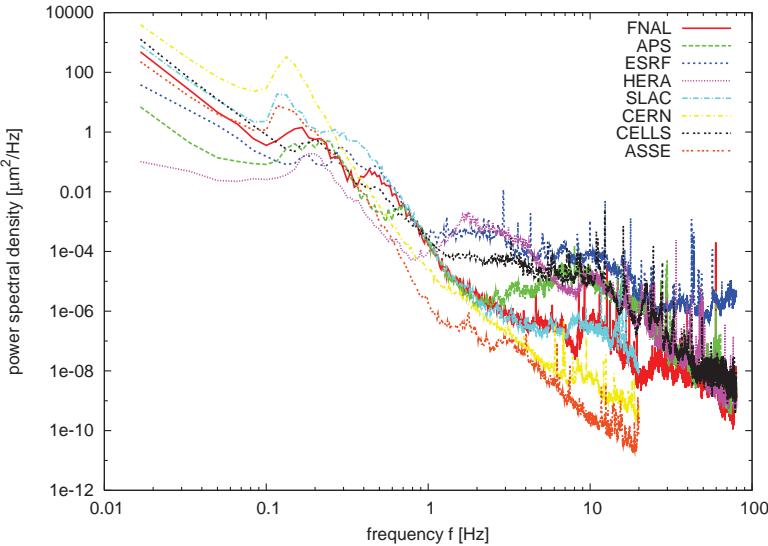


Figure 1: Example of power spectra measured at several accelerator sites [5]. An additional reference curve was taken at ASSE, a former German salt mine, about 900 m below sea level (National Datum Level –900 m).

$a = d^2z/dt^2$ . Since the charge amplifier is sensitive to a charge change  $\dot{C} = dC/dt$ , three integrations are necessary to obtain the displacement  $z$ . This results in a rather high noise level at low frequencies.

**Geophones and seismometers** Geophones basically consist of a coils suspended from a spring within the field of a permanent magnet. The voltage induced in the coil is proportional to the velocity of the coil with respect to the spring. Above the mechanical resonance frequency of typically 1 Hz, this velocity equals the instantaneous velocity of the instrument housing with respect to an inertial system. To obtain the displacement, the output signal has to be integrated once. In a seismometer, this purely passive arrangement is equipped with a PI feedback system which is designed such as to modify the mechanical transfer function of the instrument, thus resulting in a transfer function equal to that of a resonator with a resonance frequency of typical several  $10^{-3}$  Hz.

**Hydrostatic level sensors** Hydrostatic level sensors employ capacitive sensors to monitor water levels in a system of communicating tubes. Local water temperature meters are added to compensate for thermal expansion. These sensors are ideally suited for the study of spatial diffusion and correlation of ground motion.

### 5.13.4 Linacs

The rms beam displacement  $\sigma_z$  at the end of a linac due to random misalignment of all  $N_q$  quadrupoles with equal focal length  $f$  is given by

$$\sigma_z^2 = \sum_{N_q} \frac{\gamma_q}{\gamma_{\text{end}}} \bar{\beta} \bar{\beta}_{\text{end}} \frac{1}{2} \frac{\sigma_q^2}{f^2}, \quad (1)$$

with  $\bar{\beta}$  the  $\beta$ -function averaged over a FODO cell of length  $L_{\text{FODO}}$ , and the Lorentz factor  $\gamma$ . In a linac with periodic FODO lattice and  $\beta$ -function scaling  $\beta \propto \gamma^n$ , the rms beam jitter  $\sigma_z$  due to uncorrelated rms quadrupole jitter  $\sigma_q$  is

$$\sigma_z^2 = \sigma_q^2 \bar{\beta}_{\text{end}}^2 \int \frac{1}{f^2} \left( \frac{\gamma}{\gamma_{\text{end}}} \right)^{n+1} \frac{ds}{L_{\text{FODO}}}, \quad (2)$$

where the sum has been replaced by an integral, while  $2/L_{\text{FODO}}$  is the local line density of quadrupoles.

With  $ds = L_{\text{tot}} d(\gamma/\gamma_{\text{end}})$ ,  $1/f = (4/L) \sin(\mu/2)$  and  $\bar{\beta} = L/\sin \mu$  we get under the assumption of a constant phase advance  $\mu$  per FODO cell and for  $n < 0.8$  :

$$\sigma_z^2 \approx \frac{4\sigma_q^2 L_{\text{tot}}}{\bar{\beta}_{\text{end}} \sin \mu \cos^2 \frac{\mu}{2}} \frac{1}{2 - 2n}, \quad (3)$$

where  $L_{\text{tot}}$  is the linac length. With

$$N_q = \frac{2L_{\text{tot}}}{\langle L_{\text{FODO}} \rangle} \approx \frac{2L_{\text{tot}}}{(1-n)\bar{\beta}_{\text{end}} \sin \mu} \quad (4)$$

the rms beam jitter at the end of the linac is [6]

$$\sigma_z = \frac{\sigma_q \sqrt{N_q}}{\cos \frac{\mu}{2}}. \quad (5)$$

This beam motion can lead to emittance growth due to wakefield excitation [7] and to luminosity loss in colliding beam facilities. If both beams of beam size  $\sigma_{\text{beam}}$  jitter independently with rms-value  $\sigma_z$ , this results in a luminosity degradation according to

$$\mathcal{L} = \mathcal{L}_0 \exp \left( -2 \frac{\sigma_z^2}{4\sigma_{\text{beam}}^2} \right). \quad (6)$$

In a periodic FODO lattice with  $\beta \propto \sqrt{\gamma}$ , the uncorrelated quadrupole jitter tolerance can therefore be estimated at [6]

$$\begin{aligned} \sigma_q &= \sqrt{\frac{\sigma_z^2}{N_q}} \cos \frac{\mu}{2} \\ &= \sqrt{\frac{-2\sigma_{\text{beam}}^2 \ln \frac{L}{L_0}}{N_q}} \cos \frac{\mu}{2} \\ &= \sqrt{\frac{-2\epsilon_{\text{end}} \bar{\beta}_{\text{end}} \ln \frac{L}{L_0}}{N_q}} \cos \frac{\mu}{2}, \end{aligned} \quad (7)$$

with  $\epsilon_{\text{end}}$ ,  $\bar{\beta}_{\text{end}}$ ,  $N_q$ , and  $\mu$  being the geometric emittance at the end of the main linac, the average  $\beta$ -function of the last FODO cell, the number of quadrupoles per linac, and the phase advance per FODO cell, respectively.

Next we assume a single plane wave travelling at an angle  $\Theta_w$  to the linac,

$$z_{\text{gr}}(s) = \hat{z}_{\text{gr}} \cos[k_{\text{gr}} s \cos \Theta_w + \phi_0], \quad (8)$$

with  $s$  and  $\phi_0$  being the longitudinal coordinate and the wave phase, respectively.  $k_{\text{gr}}$  is the wave number of the ground motion wave. The response  $\Delta z_{\text{end}}(L)$  of the central trajectory at the end of the linac can be calculated as

$$\Delta z_{\text{end}}(L) = \int_0^L K_1(s) z_{\text{gr}}(s) R_{12}(s, L) ds. \quad (9)$$

Here  $K_1(s)$  is the quadrupole strength at position  $s$ ,  $R_{12}(s, L)$  is the (1,2)-element of the transfer matrix from  $s$  to the end of the linac,  $L$ . Assuming  $\beta \propto \sqrt{\gamma}$ ,  $K_1 \propto 1/\gamma$ , and the lengths of drift spaces and quadrupoles  $L_q \propto \sqrt{\gamma}$ , we can write [6]

$$\Delta z_{\text{end}}(L) =$$

$$(K_1 L_q)_0 \hat{z}_{\text{gr}} \sqrt{\frac{\gamma_0}{\gamma_f}} \sqrt{\beta_f} \sum_{n=0}^{N_q-1} \left( \frac{\gamma_n}{\gamma_0} \right)^{1/4} \cdot \cos(k_{\text{gr}} s_n \cos \Theta_w + \phi_0) \cdot \sqrt{\beta_0} \sin(\mu_f - n\mu/2) \frac{K_1}{|K_1|}. \quad (10)$$

Here  $s_n$  and  $\gamma_n$  are the position and energy at the  $n$ th quadrupole, while  $K_1/|K_1|$  changes sign for focusing and defocusing quadrupoles.

Time averaging yields [6]

$$\sigma_z^2 = (K_1 L_q)_0^2 \frac{\hat{z}_{\text{gr}}^2 \gamma_0}{8} \beta_f$$

$$\cdot \left| \sum_{\pm} \sum_{n=0}^{N_{\text{cell}}-1} \left( \frac{\gamma_n}{\gamma_0} \right)^{1/4} \exp[i(k_{\text{gr}} \cos \Theta_w) s_n \pm n\mu] \cdot \left( \sqrt{\hat{\beta}_0} - \sqrt{\check{\beta}_0} \exp[i(k_{\text{gr}} \cos \Theta_w L_n / 2 \pm \mu/2)] \right) \right|^2, \quad (11)$$

where  $\hat{\beta}_0$  and  $\check{\beta}_0$  are the initial maximum and minimum  $\beta$ -functions,  $N_{\text{cell}}$  is the number of FODO cells,  $L_n$  the length of the  $n$ th cell, and the sum over  $\pm$  is the sum over both the sum phase and the difference phase.

Equation 11 becomes resonant when

$$k_{\text{gr}} L_c \pm \mu = 2\pi p, \quad p \text{ integer}. \quad (12)$$

With the scalings mentioned above,  $\sigma_z^2 = \langle z_c^2 \rangle$  becomes [6]

$$\langle z_c^2 \rangle \approx$$

$$\frac{\pi N_q \tan \mu/2 \gamma_0}{2L_0^3 \gamma_f} \left( \frac{z_{\text{gr}}}{k_{\text{gr}} \cos \Theta_w} \right)^2 \cdot \sum_{\text{resonances}} (2\pi p \pm \mu) \left( \sqrt{\hat{\beta}_0} - (-1)^p \sqrt{\check{\beta}_0} \right)^2 \quad (13)$$

The assumption of correlated quadrupole motion tends to be appropriate at low frequencies. At high frequencies ground motion is uncorrelated even for short distances of a few meters and no longer wave-like.

### 5.13.5 Circular Accelerators

For uncorrelated misalignment of quadrupoles in a periodic lattice, the rms closed orbit distortion  $\sigma_{\text{co}}(s)$  at an azimuthal position  $s$  is

$$\sigma_{\text{co}} = \frac{\sqrt{\beta(s)} \sqrt{\langle \beta \rangle}}{2 \sin \pi \nu} \frac{\sigma_q}{|f|} \sqrt{N}, \quad (1)$$

with betatron tune  $\nu$ , focal length  $f$ , transverse rms quadrupole misalignment  $\sigma_q$  and number  $N$  of identical FODO cells. The average  $\langle \beta \rangle$  is taken at the quadrupoles.

Consider a storage ring of circumference  $C = 2\pi R$  which is composed of  $N$  identical FODO cells and identical bends between the quads. The betatron phase advance per FODO cell is  $\mu$ , so the tune of the ring is  $\nu = n\mu/2\pi$ .

**Vertical orbit distortion due to a single vertical ground wave** Motion of the  $n$ th quadrupole at azimuthal angle  $\Theta_n$ , measured with respect to the observation point at  $\Theta = 0$ , due to a ground wave with amplitude  $\hat{y}$ , phase  $\phi_0$  with respect to the ring center, frequency  $\omega/2\pi$ , velocity  $v$ , wavelength  $\lambda = v/(\omega/2\pi)$ , and direction of incidence  $\Theta_w$ :

$$\Delta y_n(t) =$$

$$\hat{y} \Re \exp \left\{ i \left[ \omega t + \frac{\omega R}{v} \cos(\Theta_n - \Theta_w) + \phi_0 \right] \right\}. \quad (2)$$

The closed orbit distortion at the observation point with  $\beta$ -function  $\beta_0$  due to  $\Delta y_n$  is

$$y_c = \Re \left\{ \frac{\sqrt{\beta_0}}{2 \sin \pi \nu_y f} \hat{y} \exp[i(\omega t + \phi_0)] \cdot \left[ \sum_{n=2}^{2N} \sqrt{\hat{\beta}} W(n) - \sum_{n=1}^{2N-1} \sqrt{\check{\beta}} W(n) \right] \right\}. \quad (3)$$

with

$$W(n) = \exp[i(C/\lambda) \cos(\Phi_n/\nu_y - \Theta_w)] \cos(\Phi_n - \pi \nu_y)$$

Here,  $\Phi_n$  and  $\beta_{y,n}$  are the (vertical) betatron phase advance between the observation point and the  $n$ th magnet, and the vertical  $\beta$ -function at the  $n$ th magnet, respectively.  $\hat{\beta}$  and  $\check{\beta}$  are the  $\beta$ -functions at focusing and defocusing quads. The response  $R_y$ , i.e. the ratio of the closed-orbit distortion  $\hat{y}_c$  divided by the amplitude  $\hat{y}$  of the plane ground wave, is [8, 9]

$$R_y = \frac{\hat{y}_c}{\hat{y}} = \frac{\sqrt{\beta_0}}{2f} \left( \frac{C}{\lambda} \right) \left\{ \left[ \sum_{p=-\infty}^{\infty} J_{4p} C_{4p} - J_{4p-2} C_{4p-2} \right]^2 + \left[ \sum_{p=-\infty}^{\infty} J_{4p-1} C_{4p-1} - J_{4p-3} C_{4p-3} \right]^2 \right\}^{\frac{1}{2}}, \quad (4)$$

with the Bessel function  $J_p$ , and  $C_p = \frac{(-1)^{p+1}}{\sin(\frac{\pi p}{N} - \frac{\mu}{2})}$

$$\cdot \left\{ \sqrt{\hat{\beta}} \cos \left[ p \left( \pi \frac{N+1}{N} - \Theta_w \right) - \frac{\mu}{2} \right] - \sqrt{\check{\beta}} \cos[p(\pi - \Theta_w)] \right\}.$$

$C_p$  becomes resonant for

$$|p| = |m|N \pm \nu_y. \quad (5)$$

With  $\delta\nu_y$  the distance of  $\nu_y$  from the closest integer  $[\nu_y]$ , the maxima of the resonance term

$$\left| \sin \left( \frac{\pi p}{N} - \frac{\Delta\Phi}{2} \right) \right|^{-1} \quad (6)$$

can be expressed as

$$\frac{1}{\sin \frac{\pi}{N} \delta\nu_y}, \quad (7)$$

occurring whenever

$$p_{\text{res}} = mN + [\nu_y]. \quad (8)$$

The Bessel function  $J_p$  in Eq.5.13.5 differs significantly from zero only for arguments  $C/\lambda > p$ . Since all contributions with  $p \neq p_{\text{res}}$  are suppressed, ground waves with  $C/\lambda > p_{\text{res}}$  contribute much more to  $R_y$  than those with smaller  $C/\lambda$  (see Fig.1). Thus,  $R_y$  is small for small  $C/\lambda$  and rises in a step-like manner at  $C/\lambda = [\nu_y], N - [\nu_y], N + [\nu_y], 2N - [\nu_y]$ , etc. [9].

**Horizontal orbit distortion due to a single incident plane ground wave** Horizontal waves move the magnets not only in the transverse direction, but also longitudinally, i.e. in the direction along the closed orbit. The amount of this longitudinal motion depends on the difference between the azimuthal position  $\Theta_n$  of the magnet and the direction of incidence  $\Theta_w$ . The horizontal motion of the  $n$ th quadrupole is

$$\Delta x_n(t) = \hat{x} \cos(\Theta_n - \Theta_w) \cdot \Re \exp \left\{ i \left[ \omega t + \frac{C}{\lambda} \cos(\Theta_n - \Theta_w) - \phi_0 \right] \right\}.$$

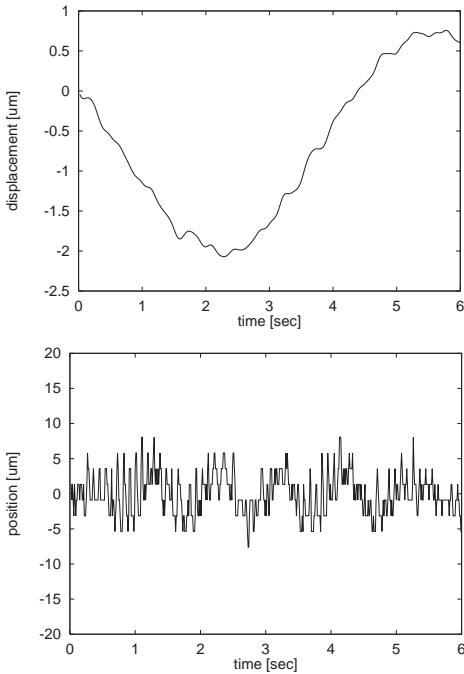


Figure 1: Example of ground motion (upper) and beam vibration (lower), normalized to  $\beta = 1 \text{ m}$ . As expected, the low frequency part of ground motion (the microseismic peak around  $1/7 \text{ Hz}$ ) is not present in the beam motion due to its long wavelength.

For the response function we get

$$\begin{aligned} R_x &= \frac{\hat{x}_c}{\hat{x}} \\ &= \frac{\sqrt{\beta_0}}{2f} \left( \frac{C}{\lambda} \right) \\ &\cdot \left\{ \left[ \sum_{p=-\infty}^{\infty} J'_{4p} C_{4p} - J'_{4p-2} C_{4p-2} \right]^2 \right. \\ &\left. + \left[ \sum_{p=-\infty}^{\infty} J'_{4p-1} C_{4p-1} - J'_{4p-3} C_{4p-3} \right]^2 \right\}^{\frac{1}{2}}, \end{aligned} \quad (9)$$

which is identical to that in the vertical case, except for replacing  $J_p$  by its derivative  $J'_p$ .

As in the case of vertical quadrupole motion, resonances occur for

$$|p| = |m|N \pm |\nu_x|, \quad (10)$$

i.e. with increasing  $C/\lambda$  the response behaves in a way similar to  $R_y$ .

**Vertical orbit distortion due to an ensemble of plane ground waves** Consider a large number  $M$  of plane ground waves of equal frequency  $\omega/2\pi$  and equal wavelength  $\lambda$ , incident from arbitrary directions  $\Theta_m$  with phases  $\phi_m$  and amplitudes  $\hat{y}_m$ . The vertical motion of the  $n$ th quadrupole is

$$\Delta y_n(t) = \sum_{m=1}^M \hat{y}_m \cdot \Re \exp \left\{ i \left[ \omega t + \frac{C}{\lambda} \cos(\Theta_n - \Theta_m) + \phi_m \right] \right\}. \quad (11)$$

The rms response  $R_y^{\text{rms}}$  yields [9]

$$\begin{aligned} R_y^{\text{rms}} &= \frac{\text{rms closed-orbit distortion}}{\text{rms ground motion}} \\ &= \frac{\sqrt{\beta_0}}{2f} \left\{ J_0^2 \left( \frac{C}{\lambda} \right) \left( \frac{\sqrt{\hat{\beta}} \cos \frac{\mu}{2} - \sqrt{\check{\beta}}}{\sin^2 \frac{\mu}{2}} \right)^2 \right. \\ &+ \sum_{p=1}^{\infty} J_p^2 \left( \frac{C}{\lambda} \right) \left[ \frac{\hat{\beta} + \check{\beta} - 2\sqrt{\hat{\beta}\check{\beta}} \cos \left( \frac{p\pi}{N} - \frac{\mu}{2} \right)}{2 \sin^2 \left( \frac{p\pi}{N} - \frac{\mu}{2} \right)} \right. \\ &\quad \left. \left. + \frac{\hat{\beta} + \check{\beta} - 2\sqrt{\hat{\beta}\check{\beta}} \cos \left( \frac{p\pi}{N} + \frac{\mu}{2} \right)}{2 \sin^2 \left( \frac{p\pi}{N} + \frac{\mu}{2} \right)} \right] \right. \\ &\quad \left. + 2 \frac{\hat{\beta} \cos \mu + \check{\beta} - 2\sqrt{\hat{\beta}\check{\beta}} \cos \frac{\mu}{2} \cos \frac{p\pi}{N}}{\cos \frac{2\pi p}{N} - \cos \mu} \right\}^{\frac{1}{2}} \end{aligned}$$

**Horizontal orbit distortion due to an ensemble of plane ground waves** The response to a horizontal compression wave can be described by the same equations as for the vertical case if one just replaces  $J_p$  by  $J'_p$  and uses the appropriate horizontal optics parameters. One must not forget an additional factor  $\sqrt{2}$  arising from averaging  $\cos^2(\Theta - \Theta_m)$  when computing the rms value,

$$\langle [\Delta x_n(t)]^2 \rangle^{\frac{1}{2}} = \left( \frac{1}{4} \sum_m \langle \hat{x}_m^2 \rangle \right)^{\frac{1}{2}}. \quad (12)$$

### 5.13.6 Numerical Modeling

While numerical modeling of ground motion in linacs is straightforward [10], simulation of ATL-like motion in circular accelerators is more involved [11].

According to the ATL model, the (vertical) position  $Y_i$  relative to a (conceptual) fixed reference point 0 after time  $T$  is given as

$$\langle Y_i^2 \rangle = ATL_{i0}, \quad (1)$$

where  $L_{i0}$  denotes the distance of location  $i$  from the fixed reference point.

From the relative motion of two locations  $i$  and  $j$

$$\langle(Y_i - Y_j)^2\rangle = \langle Y_i^2 \rangle + \langle Y_j^2 \rangle - 2\langle Y_i Y_j \rangle \quad (2)$$

it then follows that

$$\langle Y_i Y_j \rangle = AT \frac{1}{2}(L_{i0} + L_{j0} - L_{ij}) = ATM_{ij}, \quad (3)$$

or in matrix notation,

$$\langle \vec{Y} \cdot \vec{Y}^T \rangle = ATM. \quad (4)$$

$M$  is symmetric and can be diagonalized by a unitary matrix  $\Lambda$ ,

$$\tilde{M} = \Lambda \cdot M \cdot \Lambda^T, \quad (5)$$

with the components of  $\tilde{M}$  being the eigenvalues  $\lambda_i$  of  $M$ ,

$$\tilde{M}_{ij} = \lambda_i \delta_{ij}. \quad (6)$$

An appropriate choice of the reference point 0 ensures that  $\tilde{M}$  is diagonal.

Defining the vector  $\vec{V}$  as

$$\vec{V} = \Lambda \cdot \vec{Y} \quad (7)$$

and pre-multiplying Eq.4 by  $\Lambda$  and post-multiplying it by  $\Lambda^T$  yields the result

$$\langle \vec{V} \vec{V}^T \rangle = \Lambda \langle \vec{Y} \vec{Y}^T \rangle \Lambda^T = ATM \Rightarrow \langle V_i^2 \rangle = AT \lambda_i.$$

Explicitly, we construct the symmetric matrix  $M$  from the distances between the various locations under consideration and a chosen fixed reference point, and find its eigensystem. Then we generate a set of values with variances given by the eigenvalues of  $M$ , and transform these values using the eigenvectors of  $M$  to find the corresponding vertical displacements.

## References

- [1] A. Sery, O. Napoly, PRE 53 (1996)
- [2] B. Baklavov et al., Sov. Phys. Zh TF.63 (1993) 10 (in Russian)
- [3] V. Shiltsev, PRL 104, 238501 (2010)
- [4] R. Brinkmann, J. Rossbach, NIM A 350 (1994) 8-12
- [5] W. Bialowons, J. A. Osborne, and G. Shirkov, ILC-HiGrade-Report-2010-004-1
- [6] T. Raubenheimer, SLAC-387 (1991)
- [7] M. Drevlak, DESY 95-225
- [8] T. Aniel, J. L. Laclare, LNS/086 Saclay (1985)
- [9] J. Rossbach, PA 23 (1988) 121
- [10] C. Montag, DESY 97-030
- [11] A. Wolski, N. J. Walker, Proc. PAC 2003, 2396-2398 (2003)

## 5.14 VIBRATION CONTROL IN ACCELERATORS

C. Montag, BNL

(See also Sec.5.13.)

### 5.14.1 General Considerations

In the vast majority of accelerator applications, ground vibration amplitudes are well below tolerable magnet jitter amplitudes. In these cases, it is necessary and sufficient to design a rigid magnet support structure that does not amplify ground vibration. Since accelerator beam lines are typically installed at an elevation of 1 - 2 m above ground level, special care has to be taken in order to avoid designing a support structure that acts like an inverted pendulum with a low resonance frequency, resulting in intolerable lateral vibration amplitudes of the accelerator components when excited by either ambient ground motion or vibration sources within the accelerator itself, such as cooling water pumps [1] or helium flow in superconducting magnets [2]. In cases where ground motion amplitudes already exceed the required jitter tolerances, for instance in future linear colliders, passive vibration damping or active stabilization may be considered.

### 5.14.2 Passive Damping

To passively suppress the transmission of high frequency ground vibration to the accelerator magnets, a support structure with a resonance frequency  $f_r$  well below the unwanted vibration spectrum may be considered. For a single degree-of-freedom system with mass  $m$  and support spring constant  $D$ , the equation of motion can be written as

$$m\ddot{y} + k(\dot{y} - \dot{y}_0) + D(y - y_0) = 0, \quad (1)$$

with  $y(t)$  and  $y_0(t)$  being the motion of the magnet and the ground base, respectively, while  $k$  denotes the viscous damping constant. The corresponding complex transfer function denotes

$$H(s) = \frac{\omega_r^2 + 2\delta\omega_r s}{s^2 + 2\delta\omega_r s + \omega_r^2}, \quad (2)$$

with  $\omega_r = 2\pi f_r = \sqrt{D/m}$  and  $2\delta\omega_r = k/m$  and  $s = i\omega$ . According to this transfer function, a passive vibration absorber attenuates ground vibration amplitudes by a factor proportional to  $1/\omega^2$  at frequencies between the resonance frequency  $\omega_r$  and approximately  $\omega_r/\delta$ , while for even higher

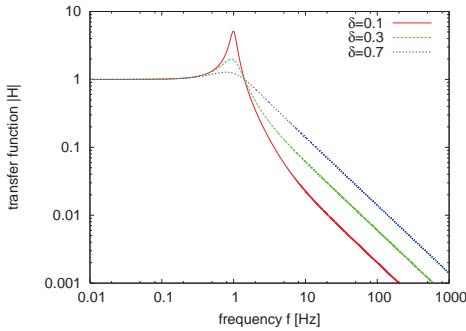


Figure 1: Transfer function  $|H|$  of a passive vibration absorber, according to Equation 2, for a resonance frequency  $f_r = 1 \text{ Hz}$  and three different damping coefficients  $\delta$ .

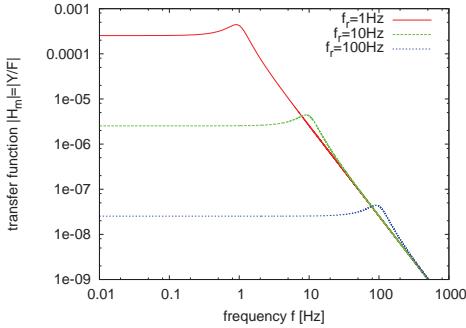


Figure 2: Transfer function  $|H_m| = |Y/F|$ , describing the displacement  $Y$  of the magnet mass as a result of frequency dependent forces  $F$  acting on the magnet itself, for a magnet mass  $m = 100 \text{ kg}$  and a damping coefficient  $\delta = 0.3$ , for three different resonance frequencies  $f_r$ . See Equation 3.

frequencies the attenuation is proportional to  $1/\omega$ , as illustrated in Figure 1.

An overshoot occurs at the resonance frequency  $f_r$ , resulting in amplification of vibration amplitudes. This effect can be eliminated by increasing the damping coefficient  $\delta$ , at the expense of a degradation of the isolation performance for high frequencies, as illustrated in Figure 1.

Though such a passive system could be designed even with a low resonance frequency of only a few Hertz, it would be very sensitive to any type of force acting on the magnet directly. Denoting the Laplace transform of the resulting

magnet vibration  $y(t)$  as  $Y(s)$ ,

$$Y(s) = \frac{\omega_r^2/D}{s^2 + 2\delta\omega_r s + \omega_r^2} F(s), \quad (3)$$

with  $F(s)$  being the Laplace transform of the input force  $f(t)$ , describes the response in the frequency domain with  $s = i\omega$ .

In the limit of low frequencies, this reduces to

$$\lim_{s \rightarrow 0} Y(s) = \lim_{s \rightarrow 0} F(s)/D, \quad (4)$$

and with  $D = \omega_r^2 \cdot m$  it becomes obvious that the sensitivity to static forces increases quadratically with decreasing resonance frequency  $f_r$ , as illustrated in Figure 2.

### 5.14.3 Active Stabilization

Stabilization of low frequencies below several tens of Hertz requires active damping. A sensor on top of the accelerator structure provides a motion signal that is processed and sent to the active support structure to counteract the detected vibration with a closed feedback loop, as schematically depicted in Figure 1. The active support can be realized as either a soft or a stiff support with an appropriate actuator.

Assuming a motion sensor with a transfer function  $H_s(s)$ , a support structure described by the (passive) transfer function

$$H_r(s) = \frac{\omega_r^2 + 2\delta\omega_r s}{s^2 + 2\delta\omega_r s + \omega_r^2}, \quad (1)$$

and a feedback algorithm  $R(s)$ , the closed-loop transfer function of the entire feedback system is computed as

$$H_g(s) = \frac{H_r(s)}{1 + R(s)H_r(s)H_s(s)}, \quad (2)$$

which describes the transfer of ground motion to the magnet.

On the other hand, the transfer function for excitation forces acting on the magnet directly becomes

$$H_m(s) = \frac{\frac{1}{D}}{1 + R(s)H_r(s)H_s(s)}. \quad (3)$$

Now, since all motion sensors (e. g. piezoelectric accelerometers, geophones, or seismometers - see section on ground vibration) have a finite lower frequency limit,  $\lim_{s \rightarrow 0} H_s(s) = 0$ , it follows that in the limit of small frequencies the active support structure exhibits the same behavior as the corresponding passive support with the same

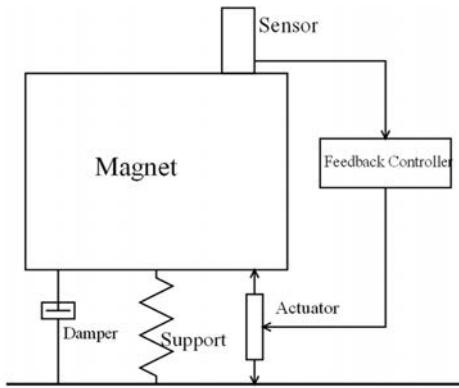


Figure 1: Schematic view of a single degree-of-freedom active stabilization system, consisting of a spring supporting the magnet, a vibration sensor to measure magnet motion, and an active feedback system to counteract these vibrations.

spring constant  $D$ , and therefore the same resonance frequency  $f_r$ ,

$$\lim_{s \rightarrow 0} H_m(s) = \frac{1}{D}. \quad (4)$$

Furthermore, since the transfer function of the support vanishes in the limit of high frequencies,  $\lim_{s \rightarrow \infty} H_r(s) = 0$ , the active support exhibits the same behavior as the passive support in the high frequency limit,

$$\lim_{s \rightarrow \infty} H_m(s) = \frac{1}{D}. \quad (5)$$

Therefore, while soft supports are efficient for vibration stabilization, static alignment of these structures is nearly impossible due to the inherent sensitivity even to small forces together with the finite lower frequency limit of the active feedback [3]. These limitations are avoided by stiff supports with resonance frequencies of hundreds of Hertz, which are typically based on piezoelectric actuators [4, 5]. For frequencies beyond 1 Hz, suppression of the rms vibration amplitude by a factor of 3 has been demonstrated.

## References

- [1] J. L. Turner, R. Stege, SLAC Collider Note 399
- [2] C. Montag et al., NIM A 564, 26-31 (2006)
- [3] S. Redaelli, CERN-AB-2004-026 (ABP), CLIC-Note-595 (2003)
- [4] C. Montag, Ph. D. thesis, DESY 97-030
- [5] C. Collette et al., NIM A (2010), doi:10.1016/j.nima.2010.05.020

# Chapter 6. ELECTRICAL CONSIDERATIONS

## 6.1 PROPERTIES OF DIELECTRICS

*M. Tigner, Cornell U.*

Table 1a) Physical attributes of some commonly used dielectrics †. Values from Mir's literature.

| Material Name                            | HA-9971            | AD99.52             | AL9953               | SAIN-174             | BeO <sup>5</sup>   | Sapphire <sup>6</sup>       | GE 2147              |
|--|--------------------|---------------------|----------------------|----------------------|--------------------|-----------------------------|----------------------|
| Purity %                                 | 99.7‡              | 99.5‡               | 99.5‡                | 99.5                 | 99.5               | ≈100,Xtal‡                  | 99.99‡‡              |
| Specific gravity                         | 3.95               | 3.89                | 3.86                 | 3.26                 | 3.0                | 3.98                        | 2.2                  |
| Flexural strength (MPa)                  | 300                | 379                 | 310                  | 300                  | 151                | 447                         | 480                  |
| Compressive str. (GPa)                   | 1.6                | 2.6                 | 2.1                  | 2.3                  | 1.6                | 2.1                         | 1.1                  |
| Young's modulus (GPa)                    | 390                | 372                 | 372                  | 270                  | 344                | 380                         | 720                  |
| Poisson's ratio                          | -                  | 0.22                | -                    | 0.25                 | 0.26               | -                           | 0.17                 |
| Therm. exp. (10 <sup>-6</sup> /K) (25°C) | 8.1                | 8.2                 | 8.3                  | 4.5                  | 9.0                | 5.01,6.7   c                | 0.55                 |
| Specific heat (J/kg·K) (25°C)            | -                  | 880                 | -                    | -                    | 1050               | 754                         | 670                  |
| Therm. cond. (W/mK) (25°C)††             | 31                 | 35.6                | 29.3                 | >170                 | 251                | 38                          | 1.4                  |
| Thermal shock res. (°C)                  | 140                | 200                 | -                    | 300                  | >1500              | -                           | 1500                 |
| Max oper temp. (°C)                      | -                  | 1750                | 1725                 | -                    | 1900               | 1900                        | 1120                 |
| Vol. resistivity (Ω·m) 25°C              | >10 <sup>12</sup>  | >10 <sup>12</sup>   | >10 <sup>12</sup>    | >10 <sup>12</sup>    | 10 <sup>13</sup>   | >10 <sup>17</sup>           | >10 <sup>17</sup>    |
| 500                                      | 10 <sup>9</sup>    | 2.1x10 <sup>8</sup> | 6x10 <sup>6</sup>    | -                    | -                  | 10 <sup>9</sup>             | 10 <sup>5</sup>      |
| 700                                      | 10 <sup>7</sup>    | 4.4x10 <sup>6</sup> | -                    | 1 x 10 <sup>5</sup>  | -                  | 10 <sup>6</sup>             | <10 <sup>4</sup>     |
| Refractive index (Na D)                  | -                  | -                   | -                    | -                    | -                  | 1.781††††                   | 1.458                |
| Transmission band (μ)                    | -                  | -                   | -                    | -                    | -                  | 0.3 - 5                     | 0.27 - 3             |
| Dielectric const. ε <sub>r</sub> ‡‡‡     | 1 MHz              | 10.0                | 9.7                  | 9.6                  | 6.5                | 9.4  , 11.5                 | 3.8                  |
| **                                       | 1 GHz              | 9.9                 | 9.7                  | 9.3                  | -                  | c-axis over freq.<br>range* | 3.8                  |
|  | 10                 | -                   | -                    | 9.4                  | 6.7                |                             | 3.8                  |
| Loss tangent (ε'"/ε') ‡‡‡                | 1 MHz              | 4x10 <sup>-5</sup>  | 2x10 <sup>-4</sup>   | 2.9x10 <sup>-4</sup> | 1x10 <sup>-3</sup> | <1x10 <sup>-4</sup>         | 1 x 10 <sup>-4</sup> |
| 1 GHz                                    | -                  | 2x10 <sup>-4</sup>  | 1.3x10 <sup>-4</sup> | -                    | -                  | -                           | 6 x 10 <sup>-5</sup> |
| 10                                       | 1x10 <sup>-4</sup> | -                   | 8.4x10 <sup>-4</sup> | 4x10 <sup>-3</sup>   | 4x10 <sup>-3</sup> | 2x10 <sup>-4</sup>          | 2 x 10 <sup>-4</sup> |
| Dielectric str. (MV/m)                   | 15                 | 17(1.3mm)           | 31                   | 1.5                  | 9.3                | 48                          | 50                   |
| Sec. emiss coeff @ 1 kV 8                | ≈5                 | -                   | -                    | ≈2                   | 3.4;2000V          | ≈7.5                        | 2.4;400V             |

\*\*here  $\epsilon = \epsilon_r \cdot \epsilon_o = (\epsilon' - \epsilon'') \cdot \epsilon_o$  † Available as rod, plate, disk and tube

†† See Fig. 1 for temp. variations \* for microwave measurements, calibrate perturber in TM01 cavity

††† ordinary ray, varies from 1.83 at 0.27 μ to 1.59 at 5.6 μ

9-98

8354248

Table 1b) Physical attributes of some commonly used dielectrics †.

Values from Mfr's. literature .

| Material Name                 | MACOR <sup>9</sup>             | PYREX <sup>10</sup>                               | VYCOR <sup>11</sup>                        | PTFE <sup>12</sup>                            | Polystyrene <sup>13</sup>  | Polyethylene <sup>14</sup>                                      | PMMA <sup>15</sup>          |
|-------------------------------|--------------------------------|---|--|---|--|---|-----------------------------|
| Specific gravity              | 2.52                           | 2.23  | 2.18                                       | 2.1   | 1.04   | 0.92  | 1.17-1.2                    |
| Flexural strength (MPa)       | 94                             | >7  | 48   | 24  | 69   | 13.8(tensile)   | 38-70 (tens.)               |
| Compressive str. (GPa)        | 0.35                           | -   | -  | .014  | .08  | -   | -                           |
| Young's modulus (GPa)         | 67                             | 62  | 66   | 0.3(tens.)                                    | 3  | 0.14(tensile)   | 2.5-3.5                     |
| Poisson's ratio               | 0.29                           | 0.2   | 0.19                                       | -   | -  | -   | 0.4                         |
| Therm. exp. (25°C) (10-6/K)   | 0.74                           | 3.3   | 0.75                                       | 71.8  | 70   | 100-200   | 70                          |
| Specific heat (25°C) (J/kg.K) | 790                            | 750   | 750  | 921   | 1200   | 2304  | 1470                        |
| Therm. cond. (25°C) (W/mK)    | 1.5                            | 1.1   | 1.4  | 0.25  | 0.1  | 0.33  | 0.19                        |
| Thermal shock res. (°C)       | -                              | 180   | 1000                                       | -   | -  | -   | -                           |
| Max oper temp. (°C)           | 500                            | 230   | 900  | 190   | 82   | 60-70   | 70                          |
| Vol. resistivity 25°C (Ω.m)   | 10 <sup>14</sup><br>500<br>700 | 10 <sup>12</sup><br>10 <sup>4</sup><br>-          | 10 <sup>10</sup><br>3x10 <sup>4</sup><br>- | 10 <sup>16</sup><br>-<br>-                    | >10 <sup>14</sup><br>-<br>-  | 10 <sup>15</sup><br>-<br>-                                      | >10 <sup>14</sup><br>-<br>- |
| Refractive index (Na D)       | -                              | 1.47  | -  | -   | 1.6  | 1.5-1.54  | 1.49                        |
| Transmission band (μm)        | -                              | 0.5 - 2.5   | 0.4 - 2.2††                                | -   | -  | -   | visible                     |
| Dielectric const. ε r         | 1 MHz<br>**<br>1 GHz<br>10 GHz | 6.0<br>-<br>-                                     | 4.6<br>-<br>-                              | 3.8<br>2.1<br>3.8                             | 2.1<br>2.56<br>2.55<br>2.1   | 2.26<br>2.26<br>2.26<br>2.26                                    | 2.7<br>-<br>-<br>-          |
| Loss tangent (ε''/ε')         | 1 MHz<br>**<br>10 MHz<br>1 GHz | 4.7x10 <sup>-3</sup><br>-<br>7.1x10 <sup>-3</sup> | 5x10 <sup>-3</sup><br>-<br>-               | 5x10 <sup>-4</sup><br>-<br>6x10 <sup>-4</sup> | <2x10 <sup>-4</sup> 2x10 <sup>-4</sup><br>6x10 <sup>-4</sup><br>1x10 <sup>-3</sup> | <2x10 <sup>-4</sup><br>3x10 <sup>-4</sup><br>6x10 <sup>-4</sup> | 0.02<br>-<br>-              |
| Dielectric str. (MV/m)        | 40                             | 250‡  | -  | 39  | 23   | 47  | 15-22                       |

\*\*here  $\epsilon = \epsilon_r \epsilon_0 = (\epsilon' - i\epsilon'') \epsilon_0$ 

† Available as rod, plate, disk and tube

‡ 1mm thk. in oil

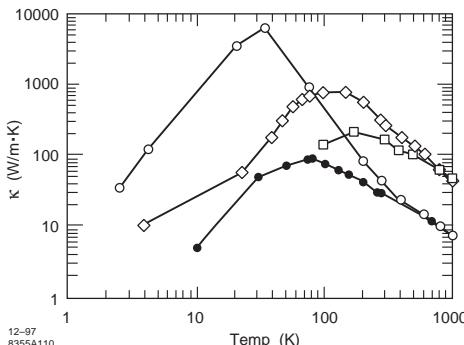


Figure 1: Thermal Conductivity of Sapphire (○), BeO(◊), Al<sub>2</sub>O<sub>3</sub> (●), AlN.

**Solid dielectrics** Consult the websites <http://www.matweb.com/index.aspx> and <http://www.omnexus.com> for further information (also on conducting material).

**Fluid dielectrics** See Sec. 7.2.9 for non-oil based fluids. For insulating transformers, rectifier banks and other off beam line devices it is common to use an oil such as Shell Diala® A or AX. Dielectric breakdown voltage, 60 Hz, for disk electrodes with 1 mm gap is typically >35 kV and under the impulse test w. needle to sphere gap of 25.4 mm breakdown should exceed 190 kV. The power factor at 60 Hz is typically 0.002. Volume coefficient of expansion is  $7 \times 10^{-4}$ °C, dielectric constant 2.2-2.3, specific heat 0.45 cal/gm, and thermal conductivity  $3 \times 10^{-4}$  cal/cm s °C.

### Notes for Tabs.1 and 2

1. Al<sub>2</sub>O<sub>3</sub>, Technical Ceramics, 75 Main St., Suite 303, Milburn, NJ 07041
2. Al<sub>2</sub>O<sub>3</sub>, Coors Ceramic Company, 600 Ninth St., Golden, CO 80401
3. Al<sub>2</sub>O<sub>3</sub>, WESGO Ceramics Div. 477 Harbor Blvd. , Belmont CA94002
4. AlN, Sumitomo Denko, SEUSA 212 308 6444
5. Thermalox 995, Brush Wellman Co, 6100 Tucson Blvd., Tucson, AZ 85706
6. General Ruby and Sapphire Corp. P.O.Box 610, New Port Richey, FL 34656-0610
7. Fused Quartz General Electric Co. Glen Allen, VA
8. S. Michizono et al, IEEE Trans. of Elec. Insul. V.28, No. 4, Aug. 1993
9. Machinable Glass Ceramic, Corning Inc, Corning NY, 14831
10. Glass Code 7740, Corning Inc., Corning, NY 14831
11. Glass Code 7913, Corning Inc., Corning, NY 14831

12. Polytrifluoroethylene, e.g. Teflon®, DuPont
13. General purpose
14. Unreinforced
15. Polymethylmethacrylate, e.g. "lucite", "perspex", "Plexiglas". Grades transmitting UV available.

## 6.2 PROPERTIES OF CONDUCTORS, NORMAL AND SUPERCONDUCTING

R.M. Scanlan, LBNL

**Normal conductors** Resistivities for some common conductors and resistive materials at 0°C are given in Tab.1 together with their temperature coefficients of resistance valid over the range ±200°C about 0°C [1],

$$\rho(T) = \rho(0)(1 + \alpha\Delta T) \quad (1)$$

Ag, Cu and Al are often used as the supporting matrix for superconductors also, carrying the whole current for a short while in the event of a quench of the SC state. Although the room temperature resistivity of these materials does not vary much with purity, the low temperature values

Table 1: <sup>[a]</sup> 80Ni-20Cr; <sup>[b]</sup> microhm-cm; <sup>[c]</sup> per 100°C.

| Mat'l          | Ag   | Cu   | Fe   | Al   | NiCr <sup>[a]</sup> |
|----------------|------|------|------|------|---------------------|
| $\rho^{[b]}$   | 1.50 | 1.56 | 8.57 | 2.5  | 108                 |
| $\alpha^{[c]}$ | 0.41 | 0.43 | 0.65 | 0.44 | 0.021               |

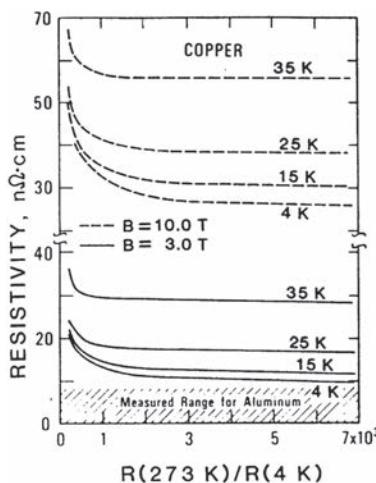


Figure 1: Resistivity as a function of purity as measured with  $B \perp I$  configuration at several magnetic fields for Cu and Al.

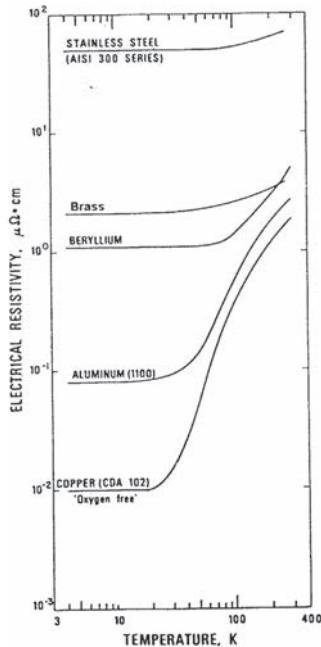
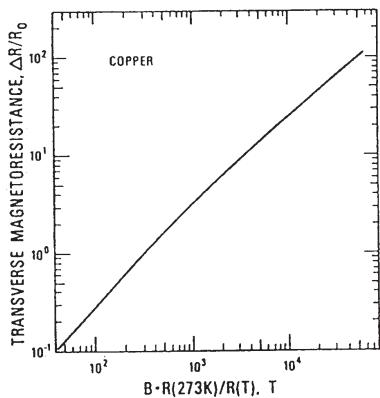
Figure 2: Resistivity of several metals vs  $T$ .

Figure 3: Kohler plot for Cu.

are strongly dependent on impurity content, the state of cold work and the ambient magnetic field. These effects [2] are illustrated in Figs.1 and 2.

The magnetoresistance effect for pure metals often<sup>1</sup> follows Kohler's rule, Fig.3,

$$\frac{\Delta\rho}{\rho} = f \left( \frac{B}{\rho(T)} \right) \quad (2)$$

<sup>1</sup>For a detailed treatment see Ref.[3]

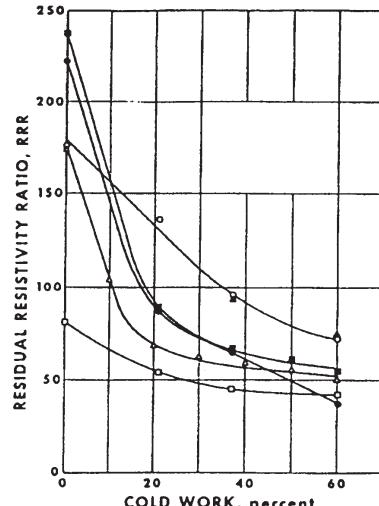


Figure 4: Effects of cold work on RRR.

Residual resistivity ratio (RRR) is a measure of purity and is defined as the ratio of the resistivity at 273K to that at 4K in the normal state. The effects of cold work of the RRR on high purity Cu and Al are shown in Fig.4.

Complete property measurements for high purity Cu can be found in [4]. The resistivity of alloys are much less dependent on temperature as seen in Tab.2.

Table 2: RRR for alloys.

| 304SS   | Invar | InconelX | 6061Al |
|---------|-------|----------|--------|
| RRR=1.4 | 1.6   | 1.05     | 2.9    |

**Superconductors** The SC elements are shown in Tab.5 in their relation to the periodic table.

Hundreds of SC compounds have been cataloged [5]. The superconducting state exists only within the boundaries of a 3-D surface in a space delimited by Temperature,  $T$ , Magnetic field,  $H$  and current density  $J$  as shown in Fig.5. The discovery of superconductivity at high temperatures in the copper oxides (1986) [6] has added hundreds more compounds to the SC material data base. Only a handful appear to possess the right combination of good intrinsic properties and fabricability that are necessary for practical devices (see Sec.6.16).

The key intrinsic properties of some practical superconductors are displayed in Tab.3. The state of the art in superconductor development is

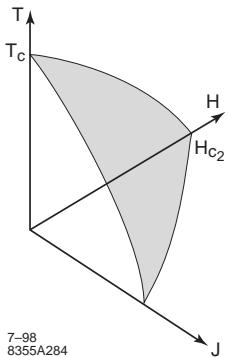


Figure 5: Superconductivity exists within the critical  $J$ - $H$ - $T$  surface.

displayed in Tab.4. Critical current density is discussed in Sec.6.16. See also Sec.7.3.10.

### Superconducting wire and cable

The most used material NbTi remains the workhorse material for most applications today with a higher field alternative, NbTiTa also available. Recent advances [7, 8] have shown the viability of Nb<sub>3</sub>Sn for some magnet applications as well. The high temperature superconductors Bi-2212, 2223 and YBCO have shown remarkable improvements and are becoming available as practical materials. Recently discovered MgB<sub>2</sub> may offer cost advantages over NbTi and Nb<sub>3</sub>Sn in certain field and temperature ranges.

Intrinsic properties of superconductors were discussed earlier. They are composition dependent but not strongly processing dependent. However, the most useful engineering property  $J_c$  depends strongly on the microstructure and processing details. The  $J_c$  of NbTi wire which has been optimized is typically 3000 A/mm<sup>2</sup> at 5 T, 4.2 K. Were the same wire to be annealed at 700–800°C for a few hours the  $J_c$  may drop by an order of magnitude, leaving  $T_c$  and  $B_{c2}$  basically unchanged. Although  $J_c$  is process dependent it is predictable and can be scaled as a function of  $T_c$  and  $H_{c2}$ . This applies to NbTi and Nb<sub>3</sub>Sn once a measurement has been made at a field and/or temperature near the proposed operating point. Eq.(3) gives the scaling rule [9] valid for  $B > 0.7 B_{c2}$ ,

$$J_c(T_0, B) = J_c(T_0, B_0) \frac{B_{c2}(T_0) - B(T_0)}{B_{c2}(T_0) - B_0(T_0)} \quad (3)$$

where  $T_0$  is the temperature at which  $J_c$  was measured,  $B_0$  is the highest magnetic induction at which the  $J_c$  is known,  $B_{c2}(T_0)$  is the critical

induction at  $T_0$  and  $J_c(T_0; B_0)$  is the critical current density at  $T_0, B_0$ .

Fig.6 [12] and Tab.4, Sec.6.2 display examples of commercially available materials as well as improvement potential. A scaling rule for Nb<sub>3</sub>Sn is also known [10] but in this case the critical values are also strain dependent [11] which needs to be taken strictly into account in practical coil design.  $J_c$  is also radiation dependent [10] increasing a factor of 2 to 3 from a fluence of 0 up to a fluence of  $10^{18}$  n/cm<sup>2</sup> and decreasing again thereafter. NbTi and Nb<sub>3</sub>Sn are available in the form of multi-filamentary composite wire and cable with a wide range of filament sizes, matrix to superconductor ratios and current ratings. For accelerator magnets where conductor magnetization is important (see Secs.7.2.2, 7.2.3)), NbTi composites with filament sizes as fine as 5 μm are standard with 2.5 μm being available on special order. The highest  $J_c$  Nb<sub>3</sub>Sn wire is limited to a filament size of 30 μm or more due to processing conditions. Cu to superconductor ratios from 1:1 to about 4:1 are easily manufacturable; higher ratios or special matrix materials such as Al for higher radiation transparency are also available on special order. Multi-kiloamp conductors are routinely available as monolithic conductors or more likely as cables of wires since it is easier to get higher current densities in small wires and cables are more flexible mechanically than monolithic conductors.

**Eddy losses in superconducting cable** Eddy currents in SC cables can cause heating and also magnetic field errors when the current and/or field conditions are changed. Intrastrand losses are reduced by twisting the strands before cabling in order to reduce the size of the eddy current loops that flow through the Cu matrix [13]. Interstrand eddy current losses can be large in Rutherford cables, especially after heat treating operations which can cause the strands to sinter together. These losses can be reduced by coating the strands with a highly resistive material and also by inserting a high resistance core material into the cable [14]. These methods can increase the interstrand resistance from ∼ 1 Ω to ∼ tens of Ω, which is adequate to reduce the eddy current coupling to a negligible level for most accelerator magnet operations.

**HTS superconductors** High current, low heat loss leads are available with HTS superconductor components, yttrium, barium copper oxide (or bismuth, strontium, calcium copper oxide) in the

## Sec.6.2: PROPERTIES OF CONDUCTORS, NORMAL AND SUPERCONDUCTING

Table 3: Superconducting properties of some selected materials at 4.2K.

| Material       | Type | $T_c$ [K] | $\mu_0 H_c$           | $\mu_0 H_{c1}$      | $\mu_0 H_{c2}$      | $\lambda$ [nm]      | Length $\xi$ [nm]  | $J_c$ [kA·mm $^{-2}$ ]      |
|----------------|------|-----------|-----------------------|---------------------|---------------------|---------------------|--------------------|-----------------------------|
| Pb             | I    | 7         | 0.0803 <sup>(a)</sup> |                     |                     | 40                  | 83                 |                             |
| Nb             | II   | 9.3       | 0.37                  | 0.25                | 0.41                | 30                  | 40                 |                             |
| In             | I    | 3.4       | 0.0281                |                     |                     |                     |                    |                             |
| Nb45-50Ti      | II   | 8.9-9.3   | 0.16                  | 0.0                 | 10.5-11.0           | 500                 | 10                 | 3(at 5 T)                   |
| $Nb_3Sn$       | 18   | 0.46      | 0.034                 | 19-25               | 200                 | 6                   | 10 (at 5 T)        |                             |
| $Nb_3Ge$       | II   | 23        | 0.16                  | 0.004               | 36-41               | 659                 | 4                  | 10 (at 5 T)                 |
| NbN            | II   | 16-18     | 0.16                  | 0.004               | 20-35               | 600                 | 5                  | 10 (at 0 T)                 |
| $PbMo_6S_8$    | I    | 14-15     | 0.4                   | 0.005               | 40-55               | 240                 | 4                  | 0.8 (at 5 T)                |
| $YBa_2Cu_3O_7$ | II   | 92        | 0.5                   | 0.05 <sup>(b)</sup> | 60 <sup>(b)</sup>   | 150 <sup>(b)</sup>  | 15 <sup>(b)</sup>  | 1(77 K, 0 T) <sup>(d)</sup> |
| Sn             | I    | 3.7       | 0.03                  |                     | >200 <sup>(c)</sup> | 1000 <sup>(c)</sup> | 2-3 <sup>(c)</sup> |                             |
| $MgB_2$        | II   | 39        |                       | 3.3 <sup>(b)</sup>  | 48 <sup>(b)</sup>   | 51 <sup>(b)</sup>   |                    |                             |
|                |      |           |                       | 14.8 <sup>(c)</sup> | 34 <sup>(c)</sup>   | 13 <sup>(c)</sup>   | 1(at 5 T)          |                             |

(a) Thermodynamic critical field at 0 K. (b) Measured with field parallel to the c-axis.

(c) Measured with field parallel to the a-b plane. (d) Epitaxial thin film, current in the a-b plane.

MF = multi-filamentary

Table 4: State of the art in superconductor development.

| Material    | Type                            | Critical Current Density [A/mm $^2$ ] | Piece Length | Manufacturer         |
|-------------|---------------------------------|---------------------------------------|--------------|----------------------|
| Nb-46.5% Ti | MF wire<br>0.1 to 2.0 mm        | 3000 [at 4.2 K, 5T]                   | 10 km        | OST, Luvata, EAS     |
|             | MF wire                         | 3000 [at 4.2 K, 12T]                  |              | Furukawa and others  |
| $Nb_3Sn$    | 0.5 to 2.0 mm<br>RRP process    | 1500 [at 4.2 K, 15T]                  | km           | OST                  |
|             | MF wire                         | 1000 [at 4.2 K, 12T]                  |              | EAS, OST, Luvata     |
| $Nb_3Sn$    | 0.5 to 2.0 mm<br>Bronze process | 750 [at 4.2 K, 15T]                   | km           | and others           |
|             | MF wire                         | 2000 [at 4.2 K, 15T]                  |              | Showa, OST           |
| Bi-2212     | 0.5 to 2.0 mm<br>PIT process    |                                       | km           |                      |
|             | Monolithic tape                 | 14000 [at 4.2 K, 15T]                 |              | Superpower, ASC, and |
| YBCO        | 4-12 mm wide<br>0.1 mm thick    |                                       | km           | others               |
|             | PIT process                     | 300 [at 4.2 K, 10T]                   |              | HyperTech, Columbus  |

range 100 to 10,000 A at a price of \$ 3/A/lead. First generation HTS wire Bi-2212[15] is available in km lengths with engineering current densities exceeding  $Nb_3Sn$  at fields above 18 T. Second generation YBCO[16] is available in several hundred meter lengths, but can be manufactured only as tapes and not round wires.

## References

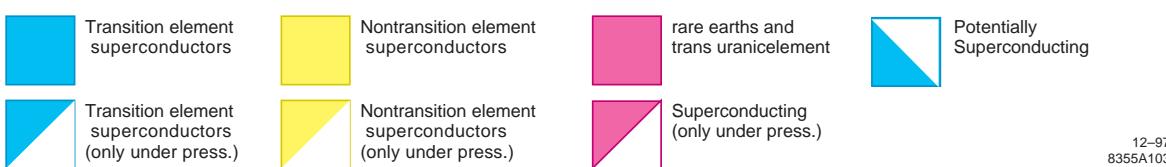
- [1] D.E. Gray, AIP Handbook, 3rd Ed. McGraw-Hill (1972) p.4
- [2] R.P. Reed, A.F.Clark, Ed's. Materials at Low Temperatures, American Society for Materials (1983)
- [3] N. Luo, G. H. Miley, Kohler's Rule and Relaxation Rates in high  $T_c$  superconductor, Physica C 371 (2002)p. 259
- [4] N.J. Simon et al, Properties of Copper and Copper Alloys at Low Temperatures NIST Monograph 177, U.S. Gov't. Printing Office (1992)
- [5] W. Warnes, Principles of Superconductivity, Metals Handbook, American Society for Materials, Vol.2 (1990) p.1030
- [6] P.F. Dahl, Superconductivity, Its Historical Roots and Development from Mercury to Ceramic Oxides, AIP (1992)
- [7] A. McInturff et al, PAC 97
- [8] R.M. Scanlan et al, 3rd European Conf. on Applied Superconductivity (1997)

| 1<br>IA                          | 2<br>IIA                        | 3<br>III                                | 4<br>IVB                           | 5<br>VB                         | 6<br>VIB                           | 7<br>VIIIB                        | 8<br>VIII                       | 9<br>VIII                      | 10<br>VIII                   | 11<br>IB                      | 12<br>IIB                   | 13<br>IIIA                    | 14<br>IVA                  | 15<br>VA                      | 16<br>VIA                    | 17<br>VIIA                   | 18<br>VIIIA                 |                            |
|----------------------------------|---------------------------------|---|------------------------------------|---------------------------------|------------------------------------|-----------------------------------|---------------------------------|--------------------------------|------------------------------|-------------------------------|-----------------------------|-------------------------------|----------------------------|-------------------------------|------------------------------|------------------------------|-----------------------------|----------------------------|
| 1 H<br>Hydrogen<br>1.00794       | 2 Be<br>Beryllium<br>9.012182   | 3 Li<br>Lithium<br>6.941                | 4 Mg<br>Magnesium<br>22.989770     | 5 Ca<br>Calcium<br>44.955910    | 6 V<br>Vanadium<br>50.9415         | 7 Cr<br>Chromium<br>51.9961       | 8 Mn<br>Manganese<br>55.845     | 9 Fe<br>Iron<br>56.8934        | 10 Co<br>Cobalt<br>58.6934   | 11 Ni<br>Nickel<br>58.6934    | 12 Cu<br>Copper<br>63.546   | 13 Zn<br>Zinc<br>65.39        | 14 Ga<br>Gallium<br>69.723 | 15 Ge<br>German.<br>67.261    | 16 As<br>Arsenic<br>74.92160 | 17 Se<br>Selenium<br>78.96   | 18 Cl<br>Chlorine<br>79.904 | 2 He<br>Helium<br>4.002602 |
| 19 K<br>Potassium<br>39.0983     | 20 Ca<br>Calcium<br>44.955910   | 21 Sc<br>Scandium<br>44.955910          | 22 Ti<br>Titanium<br>47.867        | 23 V<br>Vanadium<br>50.9415     | 24 Cr<br>Chromium<br>51.9961       | 25 Mn<br>Manganese<br>54.938049   | 26 Fe<br>Iron<br>55.845         | 27 Co<br>Cobalt<br>56.8934     | 28 Ni<br>Nickel<br>58.6934   | 29 Cu<br>Copper<br>63.546     | 30 Zn<br>Zinc<br>65.39      | 31 Ga<br>Gallium<br>69.723    | 32 Ge<br>German.<br>72.61  | 33 As<br>Arsenic<br>74.92160  | 34 Se<br>Selenium<br>78.96   | 35 Br<br>Bromine<br>79.904   | 36 Kr<br>Krypton<br>83.80   |                            |
| 37 Rb<br>Rubidium<br>85.4678     | 38 Sr<br>Strontium<br>87.62     | 39 Y<br>Yttrium<br>88.90585             | 40 Zr<br>Zirconium<br>91.224       | 41 Nb<br>Niobium<br>92.90638    | 42 Mo<br>Molybd. Technet.<br>95.94 | 43 Tc<br>Technet.<br>(97.907215)  | 44 Ru<br>Ruthen.<br>101.07      | 45 Rh<br>Rhodium<br>102.90550  | 46 Pd<br>Palladium<br>106.42 | 47 Ag<br>Silver<br>107.86882  | 48 Cd<br>Cadmium<br>112.411 | 49 In<br>Indium<br>114.818    | 50 Sn<br>Tin<br>118.710    | 51 Sb<br>Antimony<br>121.760  | 52 Te<br>Tellurium<br>127.60 | 53 I<br>Iodine<br>126.90447  | 54 Xe<br>Xenon<br>131.29    |                            |
| 55 Cs<br>Cesium<br>132.90545     | 56 Ba<br>Barium<br>137.327      | 57 La<br>Lanthan.<br>—nides<br>138.9055 | 58 Ce<br>Cerium<br>140.116         | 59 Pr<br>Praseo.<br>140.90765   | 60 Nd<br>Neodym.<br>144.24         | 61 Pm<br>Prometh.<br>(144.912745) | 62 Sm<br>Samarium<br>150.36     | 63 Eu<br>Europium<br>151.964   | 64 Gd<br>Gadolin.<br>157.25  | 65 Tb<br>Terbium<br>158.92534 | 66 Dy<br>Dyspros.<br>162.50 | 67 Ho<br>Holmium<br>164.93032 | 68 Er<br>Erbium<br>167.26  | 69 Tm<br>Thulium<br>168.93421 | 70 Yb<br>Ytterbium<br>173.04 | 71 Lu<br>Lutetium<br>174.967 |                             |                            |
| 87 Fr<br>Francium<br>(223.01731) | 88 Ra<br>Radium<br>(226.025402) | 89–103<br>Acti-<br>nides<br>(261.1089)  | 104 Rf<br>Rutherford<br>(262.1144) | 105 Db<br>Dubnium<br>(263.1186) | 106 Sg<br>Seaborg.<br>(262.1231)   | 107 Bh<br>Bohrium<br>(265.1306)   | 108 Hs<br>Hassium<br>(266.1378) | 109 Mt<br>Meitner<br>(269.273) | 110<br>(272)                 | 111<br>(277)                  | 112<br>(277)                |                               |                            |                               |                              |                              |                             |                            |

Table 5 PERIODIC TABLE OF THE ELEMENTS

|      |                               |                            |                               |                            |                                   |                             |                              |                             |                               |                             |                               |                           |                               |                              |                              |
|------|-------------------------------|----------------------------|-------------------------------|----------------------------|-----------------------------------|-----------------------------|------------------------------|-----------------------------|-------------------------------|-----------------------------|-------------------------------|---------------------------|-------------------------------|------------------------------|------------------------------|
| Lan- | 57 La<br>Lanthan.<br>138.9055 | 58 Ce<br>Cerium<br>140.116 | 59 Pr<br>Praseo.<br>140.90765 | 60 Nd<br>Neodym.<br>144.24 | 61 Pm<br>Prometh.<br>(144.912745) | 62 Sm<br>Samarium<br>150.36 | 63 Eu<br>Europium<br>151.964 | 64 Gd<br>Gadolin.<br>157.25 | 65 Tb<br>Terbium<br>158.92534 | 66 Dy<br>Dyspros.<br>162.50 | 67 Ho<br>Holmium<br>164.93032 | 68 Er<br>Erbium<br>167.26 | 69 Tm<br>Thulium<br>168.93421 | 70 Yb<br>Ytterbium<br>173.04 | 71 Lu<br>Lutetium<br>174.967 |
|------|-------------------------------|----------------------------|-------------------------------|----------------------------|-----------------------------------|-----------------------------|------------------------------|-----------------------------|-------------------------------|-----------------------------|-------------------------------|---------------------------|-------------------------------|------------------------------|------------------------------|

|                    |                                   |                                |                                  |                             |                                    |                                    |                                  |                               |                                    |                                    |                                   |                                   |                                     |                                  |                                  |
|--------------------|-----------------------------------|--------------------------------|----------------------------------|-----------------------------|------------------------------------|------------------------------------|----------------------------------|-------------------------------|------------------------------------|------------------------------------|-----------------------------------|-----------------------------------|-------------------------------------|----------------------------------|----------------------------------|
| Actinide<br>series | 89 Ac<br>Actinium<br>(227.027747) | 90 Th<br>Thorium<br>(232.0381) | 91 Pa<br>Protactin.<br>231.03588 | 92 U<br>Uranium<br>238.0289 | 93 Np<br>Neptunium<br>(237.048166) | 94 Pu<br>Plutonium<br>(244.064197) | 95 Am<br>Americ.<br>(243.061372) | 96 Cm<br>Curium<br>247.070346 | 97 Bk<br>Berkelium<br>(247.070298) | 98 Cf<br>Californ.<br>(251.079579) | 99 Es<br>Einstein.<br>(252.08297) | 100 Fm<br>Fermium<br>(257.095096) | 101 Md<br>Mendelev.<br>(258.098427) | 102 No<br>Nobelium<br>(259.1011) | 103 Lr<br>Lawrence<br>(262.1098) |
|--------------------|-----------------------------------|--------------------------------|----------------------------------|-----------------------------|------------------------------------|------------------------------------|----------------------------------|-------------------------------|------------------------------------|------------------------------------|-----------------------------------|-----------------------------------|-------------------------------------|----------------------------------|----------------------------------|



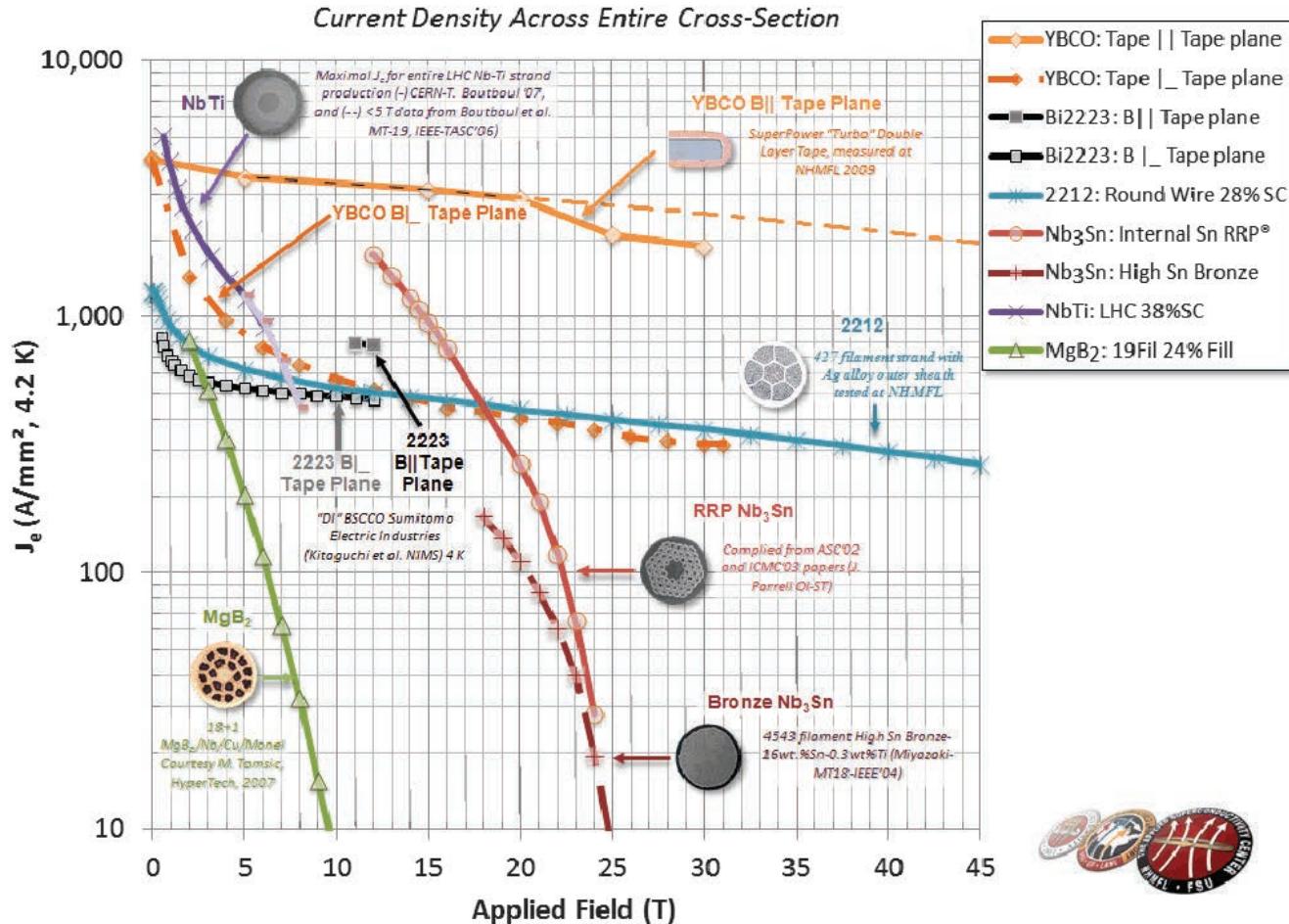


Figure 6: History of  $J_c$  improvements.

- [9] M.A. Green, IEEE Trans. on Magnetics, Vol.25, No.2 (1988)
- [10] L.T. Summers et al, IEEE Trans. on Magnetics, Vol.27, No.2 (1991)
- [11] J.W. Ekin, Strain Scaling Law for Flux Pinning in Practical Superconductors, Part 1: Basic Relationships and Applications to Nb<sub>3</sub>Sn Conductors, Cryogenics, V.20 (1980)
- [12] Courtesy P.J. Lee, U. Wisconsin
- [13] M.N. Wilson, Superconducting Magnets, Oxford (1983) p.135
- [14] M.D. Sumption et al, Adv. In Cryogenic Materials 44 (1998)
- [15] A. Goedeke, Supercond. Sci. Technol. 23 (2010) 034022
- [16] M. Rupich et al, Supercond. Sci. Technol. 23 (2010) 014015

### 6.3 PROPERTIES OF FERROMAGNETIC MATERIALS

*M.Tigner, Cornell U.*

(Secs.1.6.8, 7.2.1, 7.2.2, 7.2.3, 7.2.5, 7.2.7)

| Matl.    | $\eta$ | $x$  | $\epsilon$         |
|----------|--------|------|--------------------|
| 96Fe,Si  | 130    | 1.58 | $0.75 \times 10^7$ |
| Hipernik | 25     | 1.75 | $0.75 \times 10^7$ |

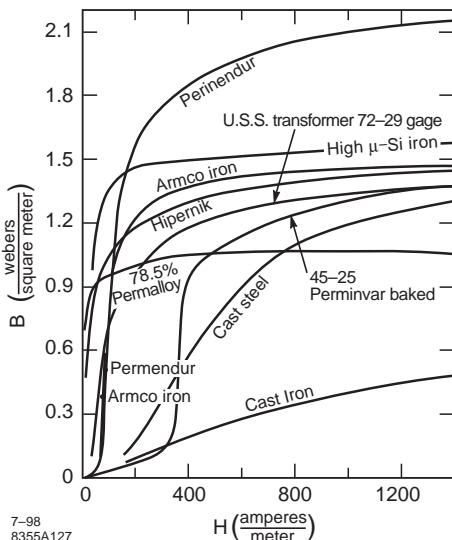


Figure 1: Magnetization curves.

**Magnet steels [1, 2, 3]** Properties of various steels are shown in Tab.1 (top, next page). AC losses in laminated steel magnets can be estimated with

$$P[W/m^3] = \eta B_m^x f + \epsilon B_m^2 d^2 f^2 \quad (1)$$

where  $f$  is the frequency,  $d$  the lamination thickness,  $B_m$  the max field in the cycle and  $\eta$ ,  $\epsilon$  and  $x$  are constants of the material. See Sec.7.2.5.

**Ferrites** The term “ferrite” is used in several ways. One class of materials, iron oxides (see Sec.6.4) are used in inexpensive permanent magnets. Another class, NiZn compounds, are used for low loss, fast response cores for pulsed magnets (Sec.7.2.7), variable permeability materials for fast tuned cavities (Sec.7.3.7) and as microwave absorbing materials (Sec.6.5)

**Permanent magnet materials** See Sec.6.4.

### References

- [1] S.S. Attwood, Electric and Magnetic Fields, Wiley (1949)
- [2] A. Goldman, Handbook of Modern Ferromagnetic Materials, Kluwer Academic Publishers, Boston/Dordrecht/London (1999)
- [3] F. Cardarelli, Materials Handbook; A Concise Desktop Reference, Springer (2008)

### 6.4 PERMANENT MAGNET MATERIALS

*R.D. Schlueter, LBNL*

(See also Sec.7.2.8.)

PMs generally used in accelerator magnets are the rare earths, NdFeB and SmCo<sub>5</sub>, Sm<sub>2</sub>Co<sub>17</sub> or ferrites, Fe<sub>2</sub>O<sub>3</sub>. The magnetic properties of these materials are described by

$$B_{\parallel} = B_r + \mu_0 \mu_{\parallel} H_{\parallel} \quad (B_{\parallel} > 0), \quad B_{\perp} = \mu_0 \mu_{\perp} H_{\perp} \quad (1)$$

where  $\parallel$  and  $\perp$  refer to the “easy” axis direction,  $\mu_0$  is the permeability of free space and  $\mu_{\parallel}$ ,  $\mu_{\perp}$  are the dimensionless relative permeabilities (see Fig.1). Typically  $(\mu_{\parallel,\perp} - 1) < 0.1$ . Uniformly magnetized PM material thus can be represented as a passive material with permeability,  $\mu_{\parallel,\perp}$  plus either active currents or charges on PM surfaces parallel or perpendicular to  $B_r$ , respectively. Magnetics effects of PM blocks are superposable to the extent that any iron in the system does not saturate. Eq.(1) describing PM material

Table 1 Magnetic properties of Steels [1]

| Composition             | $\mu_{ri}^*$ | $\mu_{rm}^*$ | $B_{sat}[T]$ | $W_\infty[J/m^3]^\dagger$ | $B_r[T]^\ddagger$ | $H_c[A/m]^{**}$ | $\rho[\Omega\cdot m]$ |
|-------------------------|--------------|--------------|--------------|---------------------------|-------------------|-----------------|-----------------------|
| 99.94Fe,C               | 250          | 5.5E3        | 2.15         | 500                       | 1.3               | 79.5            | 1E-7                  |
| 99.98Fe,C               | 25,000       | 275E3        | 2.15         | 30                        | 1.36              | 4               | 1E-7                  |
| 99.5Fe,Si               | 250          | 3.7E3        | 2.1          | 450                       | 1.28              | 63.5            | 1.8E-7                |
| 96Fe,Si§                | 400          | 6.7E3        | 2.0          | 350                       | 1.2               | 40              | 6.0E-7                |
| 50Fe,50Ni<br>"Hipernik" | 3000         | 70E3         | 1.65         | 22                        | 0.73              | 3.2             | 3.5E-7                |

\* $\mu_{ri,rm}$  are initial and maximum permeabilities relative to vacuum, †Hysteresis loss per cycle w.  
 $B_{max} = B_{sat}$ . §US Transformer 72, 29gauge .‡ $B_r$  is "residual field". \*\* $H_c$  is "coercive force".

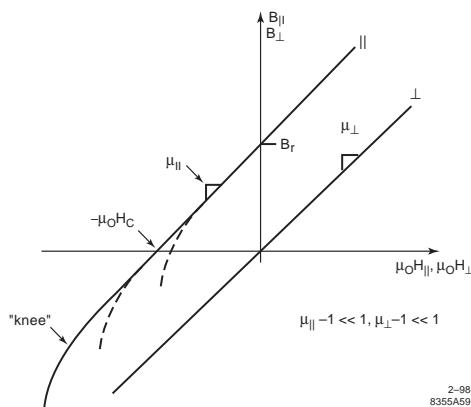


Figure 1: Permanent magnetic material  $B$ - $H$  behavior parallel and perpendicular to the magnetization direction.

behavior are valid in the second quadrant of the  $B$ - $H$  curve; linearity extends near to or somewhat into the third quadrant, where there is a "knee" in the curve. Operation beyond this knee irreversibly changes material properties and is to be avoided. One should verify that a design will not violate this condition, else actual performance may fall short of that predicted by computational code results. For some designs, optimal strength performance will be attained by selecting a more robust PM, that though having a slightly smaller  $B_r$ , has linearity extending far into the third quadrant.

NdFeB and SmCo<sub>5</sub> or Sm<sub>2</sub>Co<sub>17</sub> magnets are employed to achieve high fields in compact magnets. NdFeB PM blocks have the highest available strength, typically  $B_r = 1.1 - 1.4$  T and coercivity  $\mu_0 H_c = 1.0 - 1.3$  T. More recently, manufacturers are developing earth formulations approaching strengths of  $B_r = 1.5$  T. NdFeB is readily machinable, but its temperature stability and

radiation resistance [1, 2, 3], though often acceptable, are inferior to those of the samarium cobalt magnets. NdFeB magnets have experienced significant strength loss after exposure to neutron fluences of as low as  $10^{14}$  n/cm<sup>2</sup>. (See also Sec.3.3.9) PM materials exhibiting higher remanent fields,  $B_r$ , usually experience irreversible flux loss at lower temperatures and the knee occurs earlier in the third quadrant; thus one needs to be aware of the potential for demagnetization both in the final design configuration and during magnet assembly [4]. Typical costs are \$8/cm<sup>3</sup> for large quantities, and even higher for smaller quantities or custom-shaped blocks. Typical easy axis (i.e. magnetization vector) orientation error and block-to-block magnetization uniformity are 1° and 1%  $B_r$ , respectively. Tab.1 displays typical NdFeB properties. Major NdFeB suppliers include Hitachi Metals (formerly Sumitomo Special Metals/Neomax), Shin-Etsu, and Vacuum-schmelze.

For SmCo<sub>5</sub> or Sm<sub>2</sub>Co<sub>17</sub> one has  $B_r = 0.9 - 1.1$  T,  $\mu_0 H_c = 0.8 - 1.0$  T. It is more brittle than NdFeB, but exhibits better temperature stability and radiation resistance [5, 6]. It is thus used in more rugged thermal and radiation environments but attainable field strength is somewhat lower than for the NdFeB magnets. Typical cost is similar to or somewhat higher than that of NdFeB. Tab.1 displays typical properties of this material. Major suppliers include Shin-Etsu and Vacuum-schmelze.

Ferrites are much cheaper (\$0.02/cm<sup>3</sup>). Their remanent field and coercivity are 0.20 - 0.44 T and 0.16 - 0.35 T respectively and thus are not used for compact, high field applications. However, their low cost makes them especially suited for lower field applications. Larger, high field hybrid devices using ferrites are also feasible. Such magnets comprise the dipoles, quadrupoles and

Table 1 Typical properties of PM materials

| Property  | NdFeB          | SmCo           | $\text{Fe}_3\text{O}_4$ |
|---|----------------|----------------|-------------------------|
| $B_r$ [T]   | 1.1 -<br>1.4   | 0.9 -<br>1.2   | 0.2 -<br>0.44           |
| $\mu_0 H_c$ [T]                                     | 1.0 -<br>1.3   | 0.7 -<br>1.1   | 0.16 -<br>0.35          |
| $\mu_\perp, \mu_\parallel$                          | 1.03 -<br>1.10 | 1.01 -<br>1.03 | 1.05 -<br>1.07          |
| $\partial B_r / \partial T$ [%/deg.C]               | - 0.10         | - 0.04         | - 0.2                   |
| $\partial H_c / \partial T$ [%/deg.C]               | - 0.40         | - 0.2          | + 0.35                  |
| $T_{oper}^{\max}$ [deg. C]                          | 60 -<br>180    | >200           | 250                     |
| $T_{curie}$ [deg. C]                                | 350            | 800            | 450                     |
| $c_p$ [J/kg deg.C]                                  | 450            | 350            | 715-<br>835             |
| $\kappa$ [W/m·K]                                    | 6.4            | 10 - 23        | 4.5                     |
| $\alpha_{T,\perp}^*$ [10 <sup>-6</sup> / deg.C]     | - 0.4          | 11 - 13        | 10                      |
| $\alpha_{T,\parallel}^*$ [10 <sup>-6</sup> / deg.C] | 7              | 8 - 9          | 14                      |
| $\rho$ [g/cm <sup>3</sup> ]                         | 7.5            | 8.4            | 4.8                     |
| $\rho_r$ [10 <sup>-6</sup> Ω·cm]                    | 200            | 84             | >10 <sup>12</sup>       |

\* directional thermal expansion coefficients

combined function hybrid magnets in Fermilab's 8 GeV transfer line and recycler ring [7, 8]. These materials exhibit no magnetic degradation up to 250°C, enabling bakeouts without insulating the permanent magnet material. A disadvantage is their high temperature coefficient, - 0.20%/°C, sometimes necessitating some form of temperature compensation or control. Major suppliers include Hitachi Metals and Arnold Magnetic Technologies.

## References

- [1] J. Pfluger, G. Heintze, RSI 66-2 (1995) 1946.
- [2] O.-P. Kahkonen et al, PRB 49-9 (1994) 6052
- [3] J.R. Cost et al, IEEE Trans. Magnetics 24-3 (1988) 216
- [4] K. Robinson et al, NIM A250 (1986) 100
- [5] A.F. Zeller, J.A. Nolen, 9th In. Wkshp. Rare-Earth Mag. WP32 (1987) 157
- [6] H.B. Luna et al, NIM A285 (1989) 349
- [7] H.D. Glass et al, PAC 97
- [8] W.B. Fowler, B. Brown, J.T. Volk, PAC 97

## 6.5 PROPERTIES OF LOSSY MATERIALS

*E. Chojnacki, Cornell U.*

**General** Several types of UHV compatible microwave loads have been devised for particle accelerators, on-beam-line, off line, cryogenic, and

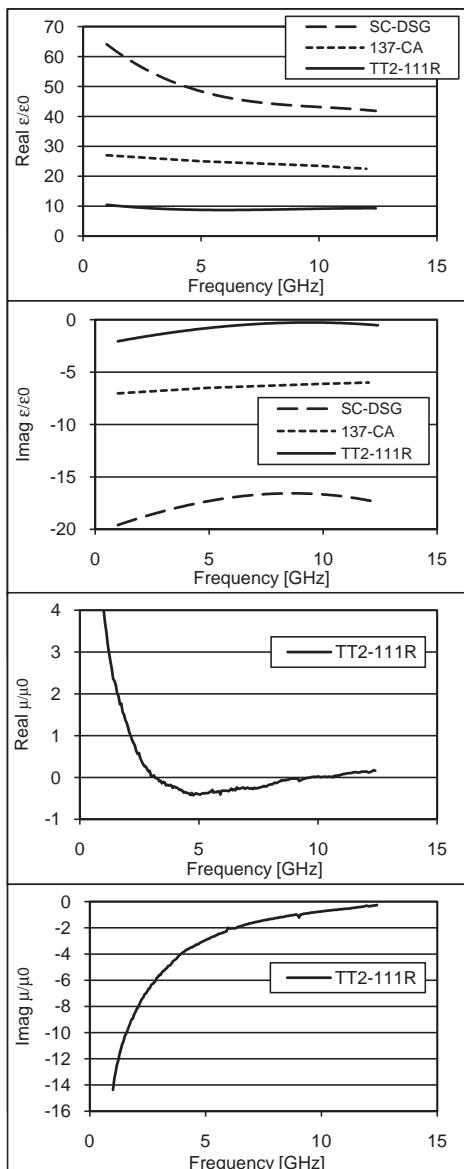


Figure 1: Permittivity and permeability of a ferrite and two artificial dielectrics at room temperature.

room temperature, with some capable of kilowatts dissipation. The most readily available robust microwave absorbing materials are ferrites and "artificial" dielectrics in which micron sized dissipative materials are dispersed within a ceramic

Table 1: Physical properties of some lossy materials. Data from mfr's literature and references.

| Name  | TT2-111R<br>[5] | C-48<br>[6] | SC-DSG (SC-35)<br>[7] | 137 CA<br>[8]  |
|---|-----------------|-------------|-----------------------|----------------|
| Composition   | Ni-Zn *         | Ni-Zn *     | SiC-C                 | AlN Composite  |
| Thermal Expansion @ 20°C ( $10^{-6}$ ) [K $^{-1}$ ] | 8.2             | 8.0         | 4.4                   | 5.0            |
| Thermal Conductivity [W/m·K]                        | 5.4             | 6.3         | 125                   | 85             |
| Spec. Heat [J/kg·K]                                 | 712             | 712         | 820                   | -              |
| Density [kg/m $^3$ ]                                | 5720            | -           | 2800                  | 2990           |
| Young's Mod. [GPa]                                  | 150             | 150         | 310                   | similar to AlN |
| Tensile Str. [MPa]                                  | 49              | 100         | -                     | -              |
| Poisson's Ratio                                     | 0.3             | 0.3         | 0.18                  | -              |
| Flex. Str. [MPa]                                    | 0.29            | 0.29        | 220                   | -              |
| Resistivity @ 20°C [ $\Omega \cdot m$ ]             | 300             | ~400        | 0.18                  | <90 @ 77K      |

\* Another similar ferrite is TDK IB004.

matrix. The most demanding application is for an on-beam-line cryogenic absorber which must have the properties of: finite DC electrical conductivity for charge drainage, low vacuum outgassing, radiation tolerant, no particulate generation, good thermal conductivity, solderable or brazable to a heat sink, and as large a microwave absorption bandwidth as possible.

**Electromagnetic properties** Fig.1 displays the permittivity and permeability of a ferrite and two artificial dielectrics at room temperature as a function of frequency. The ferrites show absorption in both electric and magnetic properties, whereas the artificial dielectrics absorb electric field energy only. These specific materials maintain their microwave loss down to 77 K, and some to 2 K. For measurement techniques, see [1, 2, 3]. The ferrites have acceptable DC conductivity for charge drainage at room temperature, but exhibit dramatic reduction in conductivity at lower temperatures [4]. The artificial dielectrics have acceptable DC conductivity if the dissipative material dispersed in the ceramic has a density above the percolation threshold, and then retains its conductivity nearly independent of temperature. The artificial dielectrics are not typically manufactured as microwave absorber products, so verification of a delivered part's electromagnetic properties must be performed by the user.

**Physical properties** Tab.1 displays physical properties of typical ferrites and artificial dielectrics.

**Heat sinking** Bonding the microwave absorber to a heat sink may require an approximate match of material thermal expansions. Tungsten

and copper-tungsten hybrids can serve as low-expansion heat sinks [9, 10]. Metallization of the microwave absorber can be a sputtered coating for soldering at 250C using eutectic Ag-Sn [9], or no metallization required for a 930C active braze using TiCuSil [11].

**Absorber assemblies** Use of these materials is illustrated by a number of UHV compatible absorber designs. Off-beam-line cavity loads using Carbon and SiC are reported in [12]-[14] while on-beam-line loads are reported in [9, 15]-[20].

## References

- [1] W. Hartung et al, CLASSE SRF Note SRF-930113/01 (1993).
- [2] W. Barry et al, LBL-33734 (1993).
- [3] R.M. Hutcheon et al, PAC 91, p. 795.
- [4] C.F. Jefferson, C.K. Baker, IEEE Trans Mag, MAG-4, No. 3, p. 460 (1968).
- [5] Product of Trans Tech Inc., 5520 Adamstown Rd, Adamstown, MD 21710.
- [6] Product of Countis Industries, 2760 Lockheed Way, Carson City, NV 89706.
- [7] Product of Coorstek Inc., 351 South Ridge Road, St. Mary's, PA 15857.
- [8] Product of Ceradyne Inc., 3169 Red Hill Ave., Costa Mesa, CA 92626.
- [9] E. Chojnacki et al, PAC 99, p. 845.
- [10] CMW Inc., 70 South Gray Street, Indianapolis, Indiana 46201.
- [11] MTC Wesgo Metals, 2425 Whipple Road, Hayward, CA 94544.
- [12] I. Campisi et al, PAC 93, p. 1220.
- [13] Rimmer et al, PAC 93, p. 819.
- [14] R. Boni et al, INFN internal note (1993).

- [15] W. Hartung et al, PAC 95, p. 3294.
- [16] S. Belomestnykh et al, PAC 95, p. 3394.
- [17] F. Hinode, S. Sakanaka, Development of a Beam Pipe HOM Absorber, ATF Damping Ring, KEK.
- [18] T. Tajima et al, PAC 95, p. 1620.
- [19] A. Novokhatski et al, Phys Rev ST Accel. Beams 10, 042003 (2007).
- [20] The European X-Ray Free-Electron Laser Technical Design Report (XFEL), DESY 2006-097, p. 70.

## 6.6 COMMON TRANSMISSION LINES AND CAVITIES

*M. Tigner, Cornell U.*

For details on the material presented here and much more the reader is referred to [1]-[6]. Herein is a collection of the formulae and mode plots often used in the accelerator field.

### Basic impedance relations

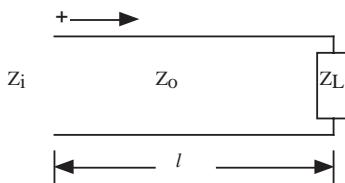


Figure 1: General line.

For a general line, Fig.1, propagating one mode and having a characteristic impedance  $Z_0$  [In Sec.6.6,  $Z_0$  is used for the characteristic impedance. The impedance of free space is designated as  $\eta_0$ .], a termination impedance  $Z_L$  and a length  $l$ , the input impedance  $Z_i$  is

$$Z_i = Z_0 \left[ \frac{Z_L \cos\left(\frac{2\pi\ell}{\lambda}\right) + iZ_0 \sin\left(\frac{2\pi\ell}{\lambda}\right)}{Z_0 \cos\left(\frac{2\pi\ell}{\lambda}\right) + iZ_L \sin\left(\frac{2\pi\ell}{\lambda}\right)} \right] \quad (1)$$

$$\Gamma = \text{reflection coefficient} = \frac{V_-}{V_+} = \frac{Z_L - Z_0}{Z_L + Z_0}$$

$$S = \text{standing wave ratio} = \frac{1 + |\Gamma|}{1 - |\Gamma|}$$

$$\hat{Z} = Z_0 S; \check{Z} = \frac{Z_0}{S}; \frac{P_{\text{refl}}}{P_{\text{inc}}} = \Gamma^2 = \left( \frac{S - 1}{S + 1} \right)^2$$

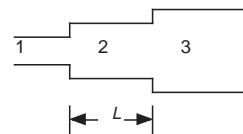


Figure 2: Quarter wave insert for matching lines. Line 1 is matched to line 3 if  $L = \lambda_g/4$  and  $Z_2 = \sqrt{Z_1 Z_3}$ .

### Surface resistance

$$\text{Skin depth } \delta_s = \sqrt{\frac{\rho}{\pi f \mu}} = \frac{0.066 \text{ [m]}}{\sqrt{f \text{ [Hz]}}} \text{ (Cu)}$$

$$R_s = \frac{\rho}{\delta_s} = 2.6 \times 10^{-7} \sqrt{f} \text{ (Cu)}$$

| mat          | Ag    | Cu  | Au       | Al <sup>[a]</sup> | W   | SS <sup>[b]</sup> | Ni <sup>[c]</sup>               |
|--------------|-------|-----|----------|-------------------|-----|-------------------|---------------------------------|
| $\rho^{[d]}$ | 1.6   | 1.7 | 2.4      | 2.8               | 5.5 | 73                | 100                             |
| [a]          | 99.6% |     |          |                   |     |                   |                                 |
| [b]          |       | 304 |          |                   |     |                   |                                 |
| [c]          |       |     | nichrome |                   |     |                   |                                 |
| [d]          |       |     |          |                   |     |                   | $10^{-8} \Omega \cdot \text{m}$ |

The anomalous skin effect When the skin depth becomes less than the mean free path of the electrons, the above relations no longer hold. In the extreme, the surface resistivity becomes independent of temperature and frequency. In the case of pure Cu at very low temperature, the improvement in dc resistivity with respect to that at room temperature might be 100 while the improvement in surface resistivity at microwave frequencies will be limited to about a factor of 6 [10].

**TEM lines** Tab.1 Shows salient properties of a number of common TEM lines.

#### Coaxial lines - additional features

$$P = \frac{V^2}{2Z_0} = \frac{E_{ri}^2 r_i^2 \sqrt{\epsilon_r}}{120} \ln\left(\frac{r_0}{r_i}\right) \quad (2)$$

$$E_r = E_{ri} \left( \frac{r_i}{r} \right); H_\phi = \frac{I}{2\pi r}; I = \frac{V}{Z_0}$$

Average power ratings for three standard rigid copper coax lines are shown in Fig.3.

Peak power rating with factor of 2 safety on the power is

$$P_{pk} = 0.06 V_{\text{test}}^2 / Z_0 \quad (3)$$

where the test voltage is such as to produce a field of about 1.5 MV/m at the inner conductor [7]. Catalog peak power ratings for 1 5/8 in, 3 1/8 in, 50Ω and 6 1/8 in, 75Ω lines are 300 kW, 1 MW and 2 MW respectively [11]. Increase factor in peak power rating in pressurized operation with dry air or SF<sub>6</sub> same as for waveguide (see later).

Table 1 Properties of TEM lines

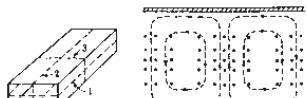
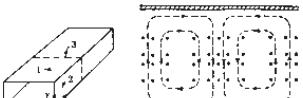
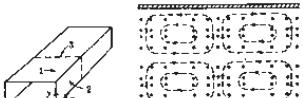
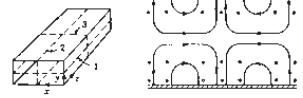
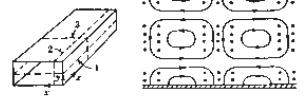
|   |   |   |   |
|---|---|---|---|
|   |   |   |   |
|   |   |   | Formulas for $a \ll b$                              |
| Capacitance, $C$<br>[farads / meter]                    | $2\pi\epsilon/\ln(r_o/r_i)$   | $\pi\epsilon/\cosh^{-1}(s/d)$   | $\epsilon b/a$                                      |
| Inductance, $L$<br>[henrys / meter]                     | $(\mu/2\pi)\ln(r_o/r_i)$  | $(\mu/\pi)\cosh^{-1}(s/d)$  | $\mu(a/b)$  |
| Conductance, $G$<br>[mhos / meter]                      | $\frac{2\pi\sigma}{\ln(r_o/r_i)} = \frac{2\pi\omega\epsilon''}{\ln(r_o/r_i)}$ | $\frac{\pi\sigma}{\cosh^{-1}(s/d)} = \frac{\pi\omega\epsilon''}{\cosh^{-1}(s/d)}$ | $\frac{\sigma b}{a} = \frac{\omega\epsilon'' b}{a}$ |
| Resistance, $R$<br>ohms / meter                         | $\frac{R_s}{2\pi} \left( \frac{1}{r_o} + \frac{1}{r_i} \right)$               | $\frac{2R_s}{\pi d} \left[ \frac{s/d}{\sqrt{(s/d)^2 - 1}} \right]$                | $\frac{2R_s}{b}$                                    |
| Characteristic<br>impedance at high<br>frequency, $Z_o$ | $\frac{\eta_o}{2\pi} \ln\left(\frac{r_o}{r_i}\right)$                         | $\frac{\eta_o}{\pi} \cosh^{-1}\left(\frac{s}{d}\right)$                           | $\eta_o \frac{a}{b}$                                |

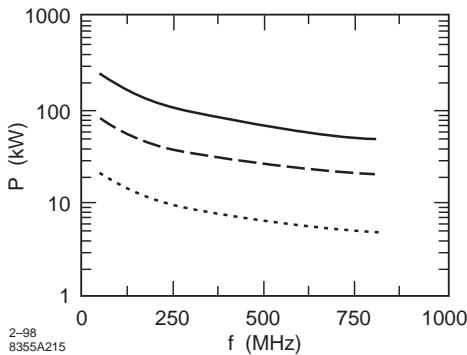
All line types  $\alpha_c$  = (atten. due to conductor) =  $R/2Z_o$ ;  $\alpha_d$  = (dielectric) =  $(\pi/\lambda)(\epsilon''/\epsilon')$

$\epsilon = \epsilon' - i\epsilon''$ ;  $\alpha_{tot}$  [db / m] =  $8.686(\alpha_c + \alpha_d)$ ; Phase constant ( $\alpha_{tot}$  small),  $\beta = \omega\sqrt{\mu\epsilon'}$

$\eta_o = \sqrt{\mu/\epsilon}$ ;  $\lambda$  = wavelength in dielectric

Table 2 Mode Patterns in Rectangular Waveguide.

| TE <sub>10</sub>  | TE <sub>11</sub>  | TE <sub>21</sub>   |
|---|---|--|
| <br>3<br>2  | <br>3<br>2  | <br>3<br>2  |
| TE <sub>20</sub>  | TM <sub>11</sub>  | TM <sub>21</sub>   |
| <br>3<br>2 | <br>3<br>2 | <br>3<br>2 |



### TM and TE lines, waveguides

General wave relations All field components are understood to have the common factor  $\exp(i\omega t - \gamma z)$  for waves traveling in the positive direction of  $z$  where  $\gamma = \text{propagation constant} = \alpha + i\beta$ ;  $\alpha = \text{attenuation constant}$ ;  $\beta = \text{phase constant} \beta = \omega\sqrt{\mu\epsilon} = 2\pi/\lambda$ ;  $v = 1/\sqrt{\mu\epsilon}$ ;  $v_g = d\omega/d\beta$ . For low loss lines, i.e.  $\alpha \approx 0$ ,  $\omega$  above cut-off:

$$\gamma = i\frac{\omega}{v}\sqrt{1 - \left(\frac{\omega_c}{\omega}\right)^2} \quad (4)$$

$$v_p = \frac{v}{\sqrt{1 - \left(\frac{\omega_c}{\omega}\right)^2}}, \quad v_p v_g = v^2 \quad (5)$$

Below cutoff,

$$\gamma = \frac{2\pi}{\lambda_c}\sqrt{1 - \left(\frac{\omega}{\omega_c}\right)^2} \quad (6)$$

the phase velocity being infinite, i.e. same phase at every  $z$ . For negatively traveling waves reverse sign of  $z$  term and of  $Z_{TE,TM}$ .

Rectangular waveguides  $TE_{m,n}$ ,  $TM_{m,n}$ :

$$f_c = \frac{1}{2\sqrt{\mu\epsilon}} \left[ \left(\frac{m}{a}\right)^2 + \left(\frac{n}{b}\right)^2 \right]^{1/2}, \quad \eta_0 = \sqrt{\frac{\mu_0}{\epsilon_0}} \quad (7)$$

$$\alpha_{TM_{m,n}} = \frac{2R_s}{b\eta_0\sqrt{1 - \left(\frac{\omega_c}{\omega}\right)^2}} \left[ \frac{m^2 \left(\frac{b}{a}\right)^3 + n^2}{m^2 \left(\frac{b}{a}\right)^2 + n^2} \right] \quad (8)$$

$$\alpha_{TE_{m,n}} = \frac{2R_s/b\eta_0}{\sqrt{1 - \left(\frac{\omega_c}{\omega}\right)^2}} \left[ \left(1 + \frac{b}{a}\right) \left(\frac{\omega_c}{\omega}\right)^2 + F \right] \quad (9)$$

$$F = \left[ 1 - \left(\frac{\omega_c}{\omega}\right)^2 \right] \left[ \frac{\frac{b}{a} \left(\frac{b}{a} m^2 + n^2\right)}{b^2 \frac{m^2}{a^2} + n^2} \right] \quad (10)$$

$TM_{m,n}$  fields: (propagating factor understood)

$$k_x = \frac{m\pi}{a}; \quad k_y = \frac{n\pi}{b}; \quad k_c^2 = k_x^2 + k_y^2$$

$$E_z = A \sin(k_x x) \sin(k_y y)$$

$$H_x = i \frac{k_y \omega}{k_c \eta_0 \omega_c} A \sin(k_x x) \cos(k_y y)$$

$$H_y = -i \frac{k_x \omega}{k_c \eta_0 \omega_c} A \cos(k_x x) \sin(k_y y)$$

$$E_x = Z_{TM} H_y; \quad E_y = -Z_{TM} H_z$$

$$Z_{TM} = \eta_0 \left[ 1 - \left(\frac{\omega_c}{\omega}\right)^2 \right]^{1/2}$$

Mode patterns are shown in Tab.2.

$TE_{m,n}$  fields:

$$H_z = B \cos(k_x x) \cos(k_y y)$$

$$E_x = i \frac{\eta_0 k_y \omega}{k_c \omega_c} B \cos(k_x x) \sin(k_y y)$$

$$E_y = -i \frac{\eta_0 k_x \omega}{k_c \omega_c} B \sin(k_x x) \cos(k_y y)$$

$$H_x = -\frac{E_y}{Z_{TE}}; \quad H_y = \frac{E_x}{Z_{TE}}$$

$$Z_{TE} = \eta_0 \left[ 1 - \left(\frac{\omega_c}{\omega}\right)^2 \right]^{1/2}$$

American Standard (E.I.A.) waveguides have designations of the form WR2100, WR1800, etc. in which the number is the inside dimension of the long side in inches  $\times$  100. The ratio of the sides is (almost always) 2:1, i.e. the short, inside dimension of WR2100 is 10.5 in.. Standard, commercially available waveguides are WR: 2300, 2100, 1800, 1500, 1150, 975, 770, 650, 510, 430, 340, 284\*, 229, 187\*, 159, 137\*, 112 \*, 90\*, 75, 62. The \* indicates some departure from 2:1.

The peak power capacity is usually taken [3] as corresponding to a maximum electric field of 2.9 MV/m at atmospheric pressure and in dry air. For 2:1 waveguides at  $\sqrt{2}\omega_c$  then

$$P = \frac{\lambda^2}{16\sqrt{2}\eta_0} (2.9 \times 10^6)^2 \quad [\text{TE}_{10} \text{ mode}] \quad (11)$$

This capacity is enhanced by pressurizing with dry air at a pressure  $P_r$  in Atmospheres as

$$\log_{10} \left( \frac{P'}{P} \right) \approx 1.24 \log_{10} P_r; \quad 1 \leq P_r \leq 6 \text{ atm} \quad (12)$$

In SF<sub>6</sub> the pressure enhancement of peak power is

$$\log_{10} \left( \frac{P'}{P} \right) = 1.87 \log_{10} P_r; \quad 1 \leq P_r \leq 2 \text{ atm} \quad (13)$$

The enhancement saturates at factor of 20 at  $P_r$  of 3 atm.

The average power handling capacity [3] is given approximately by the expression (for copper guides at a temperature rise of 42°C in free air),

$$\log_{10} P \approx -2.2 \log_{10} f [\text{GHz}] + 5.4 \quad (14)$$

provided that one operates at a frequency approximately 1.41 times the cutoff frequency.

These limits stand in contrast to the limit in vacuum, often referred to as the Kilpatrick criterion. It is an empirical relation derived from data taken before the era of UHV technology. The max. field  $E_k$  [MV/m] at any frequency  $f$  [Hz] according to this criterion is

$$f = 1.64 \times 10^6 E_k^2 \exp(-8.5/E_k) \quad (15)$$

Today, with UHV technology, fields of  $2E_k$  or more are often achieved (see Secs.6.13, 7.3.5).

Circular waveguides Mode patterns are shown in Tab.3.

Radius =  $a$ , polar coordinates  $r, \phi$ , propagation in the  $z$  direction:

$$\begin{aligned} \omega_{c,\text{TM}_{n,l}} &= \frac{p_{n,l}}{a\sqrt{\mu\epsilon}}; \quad \omega_{c,\text{TE}_{n,l}} = \frac{p'_{n,l}}{a\sqrt{\mu\epsilon'}} \\ \alpha_{\text{TM}_{n,l}} &= \frac{R_s}{a\eta_0} \frac{1}{\sqrt{1 - (\frac{\omega_c}{\omega})^2}} \\ \alpha_{\text{TE}_{n,l}} &= \alpha_{\text{TM}_{n,l}} \left[ \left( \frac{\omega_c}{\omega} \right)^2 + \frac{n^2}{p'_{n,l} - n^2} \right] \end{aligned}$$

$p$  roots of  $J_n$ ,  $p'$  roots of  $J'_n$  [8].

| Mode | $p_{n,l}$ | Mode | $p'_{n,l}$ |
|------|-----------|------|------------|
| TM01 | 2.404     | TE01 | 3.832      |
| TM02 | 5.520     | TE02 | 7.016      |
| TM03 | 8.654     | TE03 | 10.173     |
| TM11 | 3.832     | TE11 | 1.841      |
| TM12 | 7.016     | TE12 | 5.331      |
| TM13 | 10.173    | TE13 | 8.536      |
| TM21 | 5.136     | TE21 | 3.054      |
| TM22 | 8.417     | TE22 | 6.706      |
| TM23 | 11.620    | TE23 | 9.969      |

TM waves:  $Z_{\text{TM,TE}}$ ,  $\gamma$ , as before.

$$E_z = AJ_n(k_cr) \begin{cases} \cos(n\phi) \\ \sin(n\phi) \end{cases}; \quad k_c = \frac{p_{nl}}{a}$$

$$H_r = -i \frac{n\omega}{k_c \eta_0 r \omega_c} AJ_n(k_cr) \begin{cases} \sin(n\phi) \\ -\cos(n\phi) \end{cases}$$

$$H_\phi = -i \frac{\omega}{\omega_c \eta_0} AJ'_n(k_cr) \begin{cases} \cos(n\phi) \\ \sin(n\phi) \end{cases}$$

$$E_\phi = -H_r Z_{\text{TM}}; \quad E_r = H_\phi Z_{\text{TM}}$$

TE waves:

$$H_z = BJ_n(k_cr) \begin{cases} \cos(n\phi) \\ \sin(n\phi) \end{cases}; \quad k_{c_nl} = \frac{p'_{nl}}{a}$$

$$E_r = i \frac{n\eta_0 \omega}{k_c r \omega_c} BJ_n(k_cr) \begin{cases} \sin(n\phi) \\ -\cos(n\phi) \end{cases}$$

$$E_\phi = i \eta_0 \frac{\omega}{\omega_c} BJ'_n(k_cr) \begin{cases} \cos(n\phi) \\ \sin(n\phi) \end{cases}$$

$$H_\phi = \frac{E_r}{Z_{\text{TE}}}; \quad H_r = -\frac{E_\phi}{Z_{\text{TE}}}$$

Coaxial waveguides Some mode patterns for TE and TM waves in coaxial guides are shown in Figs.4 and 5. Cutoff wavelengths for the lower modes are plotted in Fig.6 [9].

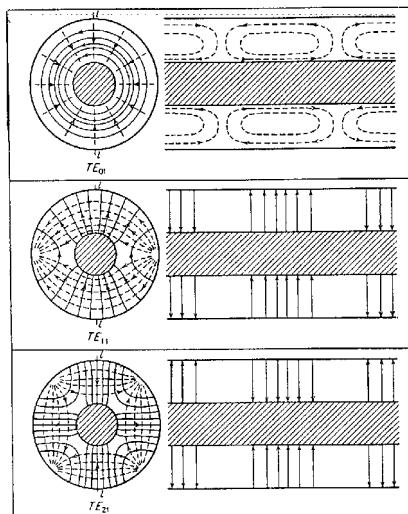


Figure 4: TE mode patterns in coax.

### Cavities

(See Sec.7.3 for applications to acceleration.)

Mode fields in cavities can be derived from a standing wave pattern of a corresponding waveguide mode. For negative traveling waves, the

Table 3 Mode Patterns in Circular Waveguide.

| Wave Type   | TM <sub>01</sub>                                 | TM <sub>02</sub>                                 | TM <sub>11</sub>   | TE <sub>01</sub>                                 | TE <sub>11</sub>   |
|---|--|--|--|--|--|
| Field distributions in cross-sectional plane, at plane of maximum transverse fields |  |  |  |  |  |
| Field distributions along guide   |  |  |  |  |  |
| Field components present  | E <sub>z</sub> , E <sub>r</sub> , H <sub>φ</sub> | E <sub>z</sub> , E <sub>r</sub> , H <sub>φ</sub> | E <sub>z</sub> , E <sub>r</sub> , E <sub>φ</sub> , H <sub>r</sub> , H <sub>φ</sub> | H <sub>z</sub> , H <sub>r</sub> , E <sub>φ</sub> | H <sub>z</sub> , H <sub>r</sub> , H <sub>φ</sub> , E <sub>r</sub> , E <sub>φ</sub> |

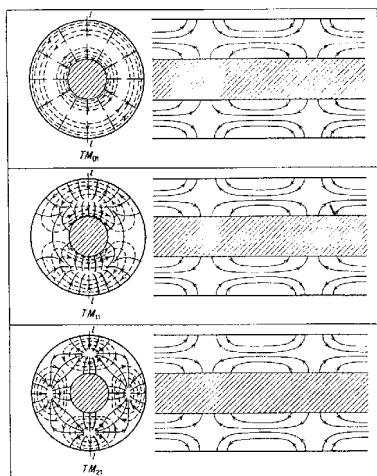
7-98  
8355A213

Figure 5: TM mode patterns in coax.

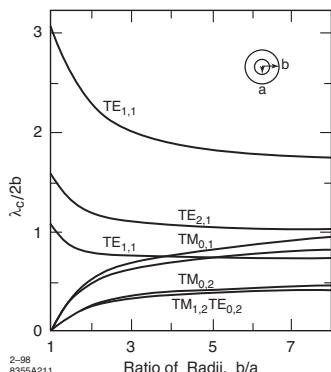


Figure 6: Cutoff wavelengths in coax.

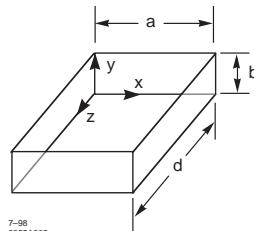


Figure 7: Rectangular cavity dimensions.

propagation factor becomes  $\exp(i\omega t + \gamma z)$  and the terms involving  $Z_{TE}$  and  $Z_{TM}$  are reversed in sign.

Rectangular cavities (Fig.7)

$TE_{101}$  mode corresponds to  $TE_{10}$  standing wave mode in a rectangular waveguide

$$E_y = E_0 \sin\left(\frac{\pi x}{a}\right) \sin\left(\frac{\pi z}{d}\right)$$

$$H_x = -i \frac{E_0}{\eta_0} \frac{\lambda}{2d} \sin\left(\frac{\pi x}{a}\right) \sin\left(\frac{\pi z}{d}\right)$$

$$H_z = i \frac{E_0}{\eta_0} \frac{\lambda}{2a} \cos\left(\frac{\pi x}{a}\right) \sin\left(\frac{\pi z}{d}\right)$$

$$U = \text{stored energy} = U_E = \epsilon' \frac{abd}{8} E_0^2$$

$$W_L = \text{wall power loss} = \frac{R_s}{2} \oint_{\text{innersurf.}} H_{\text{surf}}^2 dS$$

$$W_L = \frac{R_s \lambda^2}{8 \eta_0^2} E_0^2 \left[ \frac{ab}{d^2} + \frac{bd}{a^2} + \frac{1}{2} \left( \frac{a}{d} + \frac{d}{a} \right) \right]$$

$$Q_0 \equiv \frac{\omega_0 U}{W_L} = \frac{\pi \eta_0}{4 R_s} \left[ \frac{2b(a^2+d^2)^{3/2}}{ad(a^2+d^2)+2b(a^3+d^3)} \right]$$

$$G^{-1} \equiv \frac{(E_0 b)^2}{2 W_L} = \frac{\eta_0^2}{R_s} \left[ \frac{2b^2(a^2+d^2)}{2b(a^3+d^3)+ad(a^2+d^2)} \right]$$

Generally, the TE<sub>mnp</sub> mode fields are

$$\begin{aligned} H_x &= -\frac{C}{k_c^2} \frac{p\pi}{d} \frac{m\pi}{a} \sin(k_x x) \cos(k_y y) \cos(k_z z) \\ H_y &= -\frac{C}{k_c^2} \frac{p\pi}{d} \frac{n\pi}{b} \cos \cdot \sin \cdot \cos \\ E_x &= i \frac{\omega \mu C}{k_c^2} \left( \frac{n\pi}{b} \right) \cos \cdot \sin \cdot \sin \\ E_y &= -i \frac{\omega \mu C}{k_c^2} \left( \frac{m\pi}{a} \right) \sin \cdot \cos \cdot \sin \end{aligned}$$

and for TM<sub>mnp</sub>,

$$\begin{aligned} E_z &= D \sin(k_x x) \sin(k_y y) \cos(k_z z) \\ E_x &= -\frac{D}{k_c^2} \left( \frac{p\pi}{d} \right) \left( \frac{m\pi}{a} \right) \cos \cdot \sin \cdot \sin \\ E_y &= -\frac{D}{k_c^2} \left( \frac{p\pi}{d} \right) \left( \frac{n\pi}{b} \right) \sin \cdot \cos \cdot \sin \\ H_x &= i \frac{\omega \epsilon D}{k_c^2} \left( \frac{n\pi}{b} \right) \sin \cdot \cos \cdot \cos \\ H_y &= -i \frac{\omega \epsilon D}{k_c^2} \left( \frac{m\pi}{a} \right) \cos \cdot \sin \cdot \cos \end{aligned}$$

For both TE and TM modes the cut-off parameter, phase constant and resonant wavelength are

$$\begin{aligned} k_c^2 &= \left( \frac{m\pi}{a} \right)^2 + \left( \frac{n\pi}{b} \right)^2 \\ \beta &= \left[ \left( \frac{2\pi}{\lambda^2} \right) - k_c^2 \right]^{\frac{1}{2}} = \frac{p\pi}{d} \\ k &= \frac{2\pi}{\lambda_{\text{res}}} = \left[ \left( \frac{m\pi}{a} \right)^2 + \left( \frac{n\pi}{b} \right)^2 + \left( \frac{p\pi}{d} \right)^2 \right]^{\frac{1}{2}} \end{aligned}$$

Note that TE and TM with same  $mnp$  are degenerate. If  $a = b = d$ , 112, 121, 211 of both TM and TE are degenerate.

Cylindrical cavities Fig.8 shows the basic cylindrical or "pill box" cavity.

Coordinates  $r, \phi, z$ . (Note that it is common in circular waveguides to give the mode indices in order  $\phi, r, z$ .) One way of deriving the fields of the TM<sub>010</sub> cavity mode is from a standing

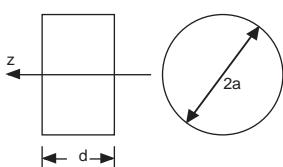


Figure 8: Pill box cavity.

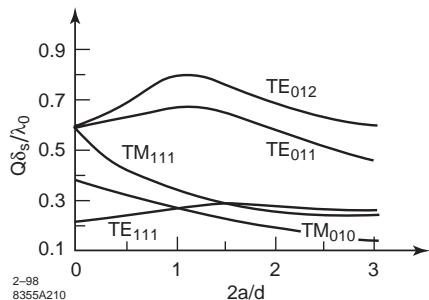


Figure 9: Mode Q's for the cylinder.

wave corresponding to the TM<sub>01</sub> mode in circular waveguide.

$$\begin{aligned} E_z &= E_0 J_0(kr); \quad k = \frac{2.405}{a}; \quad \lambda_{\text{res}} = 2.61 \text{ a} \\ H_\phi &= i \frac{E_0}{\eta_0} J_1(kr); \quad U = \pi \epsilon' d \frac{a^2}{2} J_1^2(ka) \\ Q_0 &= \frac{\eta_0}{R_s} \frac{2.405}{2 \left( \frac{a}{d} + 1 \right)}; \quad G^{-1} = \frac{\eta_0^2}{R_s} \frac{d}{2\pi a} \frac{J_1^{-2}(2.405)}{1 + \frac{a}{d}} \end{aligned}$$

For higher modes the resonant frequencies may be obtained from

$$f_{nml} = \left[ \left( \frac{x_{nm}}{a} \right)^2 + \left( \frac{l\pi}{d} \right)^2 \right]^{\frac{1}{2}} \frac{c}{2\pi} \quad (16)$$

$x_{nm} = p_{nm}$  for TM, and  $p'_{nm}$  for TE modes.

Field equations for higher TM like modes will be found in Sec.7.3.8.1. Fig.9 [6] displays the Q's for several modes.

Spherical cavities Coordinate order  $r, \phi, \theta$ . Sphere of radius  $a$ . For the TM<sub>101</sub> mode,

$$\begin{aligned} H_\phi &= \frac{C \sin \vartheta}{kr} \left( \frac{\sin kr}{kr} - \cos kr \right) \\ E_r &= -2i \frac{\eta_0 C \cos \vartheta}{k^2 r^2} \left( \frac{\sin kr}{kr} - \cos kr \right) \\ E_\theta &= i \eta_0 \frac{C \sin \vartheta}{k^2 r^2} \left[ \frac{(kr)^2 - 1}{kr} \sin kr + \cos kr \right] \end{aligned}$$

Resonance condition is found from the requirement that  $E_\theta = 0$  at  $r = a$ :  $\lambda \approx 2.29 a$  and  $Q \approx \eta_0 / R_s$ . The fields for the TE<sub>101</sub> mode are found by substituting  $E_\phi$  for  $H_\phi$ ,  $-H_r$  for  $E_r$  and  $-H_\theta$  for  $E_\theta$  in the above. The resonance condition is that  $E_\phi = 0$  at  $r = a$ , or,  $\lambda \approx 1.395 a$ .

Cavity Tuning For precise tuning of cavity frequencies it is sometimes necessary to take account of the difference of the relative dielectric

constant of air from that of vacuum owing to the finite pressure and presence of water vapor [12],

$$\frac{\Delta f}{f} = -\frac{1}{2} \frac{\epsilon_r(\text{air}) - \epsilon(\text{vac})}{\epsilon(\text{vac})}$$

$$\epsilon_{\text{air}} = 1 + 2.10 \times 10^{-4} \frac{P_a}{T}$$

$$+ 1.80 \times 10^{-4} \left( 1 + \frac{5580}{T} \right) \left( \frac{P_w}{T} \right)$$

where  $P_a$  and  $P_w$  are the partial pressures of air and water in mmHg and  $T$  the temperature in K.

**Frequency bands** Microwave frequency bands are sometimes designated with letters following older radar practice (see IEEE Standard 521-1984):

|                |             |
|----------------|-------------|
| HF             | 3-30 MHz    |
| VHF            | 30-300 MHz  |
| UHF            | 0.3-1 GHz   |
| L              | 1-2 GHz     |
| S              | 2-4 GHz     |
| C              | 4-8 GHz     |
| X              | 8-12 GHz    |
| K <sub>u</sub> | 12-18 GHz   |
| K              | 18-27 GHz   |
| K <sub>a</sub> | 27-40 GHz   |
| V              | 40-75 GHz   |
| W              | 75-110 GHz  |
| mm             | 110-300 GHz |

Sometimes the designation P is used for the band 0.42-0.45 GHz. The SLAC linac S-band is 2856 MHz ( $\lambda_{rf} = 10.5$  cm).

## References

- [1] N. Marcuvitz, Waveguide Handbook, M.I.T. Radiation Lab. Series No.10, Boston Tech Pub. (1964)
- [2] G.L. Ragan, Microwave Transmission Circuits, M.I.T. Radiation Lab. Series No.9, McGraw-Hill (1948)
- [3] T.S. Saad Ed, Microwave Engineer's Handbook, Vol.I & II Artech House (1971)
- [4] E. Ginzton, Microwave Measurements, McGraw-Hill (1957)
- [5] S. Ramo et al, Fields and Waves in Communication Electronics, Wiley (1965)
- [6] R.E. Collin, Foundations for Microwave Engineering, McGraw-Hill (1966)
- [7] Data from Andrew Corp.
- [8] E. Jahnke, F. Emde, Tables of Functions with Formulae and Curves, Dover (1945)

- [9] T. Moreno, Microwave Transmission Design Data, McGraw-Hill (1948) p.71
- [10] H. Padamsee, J. Knobloch and T. Hayes in RF Superconductivity for Accelerators, 2nd ed, Wiley - VCH p. 79
- [11] Dielectric Communications, Raymond, ME
- [12] Reference Data for Radio Engineers, 5th ed. H.W. Sams (1968)

## 6.7 RF PULSE COMPRESSION

### 6.7.1 Passive Pulse Compression

*Z.D. Farkas, SLAC*

High acceleration gradients require high peak powers. One possible way to generate high peak powers is to generate a relatively long pulse at a relatively low power and compress it into a shorter pulse with higher peak power. It is possible to compress prior to dc-to-rf conversion or at radio frequencies after dc-to-rf conversion. Three methods of rf pulse compression (RFPC) are considered: The SLAC Energy Development (SLED) [1], where rf energy is stored in rf cavities; SLED-II [2], where the energy is stored in long resonant delay lines; and the Binary Pulse Compressor (BPC) [3], where the energy is stored in traveling wave delay lines.

In all three methods, a high power amplifier, generally a klystron, is interposed between the active control element and the input to the pulse compression (PC) network, so that the active control element operates at low power. PC using switches operating at the high power levels of the amplifier output will not be considered. The output of the PC network is used as the input to an accelerator section in order to increase the section voltage (beam energy). The advantages of RFPC are:

- (i) RFPC matches the modulator pulse width of several microseconds to the accelerating structure fill time of fractions of microseconds;
- (ii) Generally the higher the power the harder it is to increase it further. This is not the case with RFPC because the control elements operate at low power;
- (iii) With SLED we can also operate with uncompressed pulses by turning the modulation off and detuning the cavities. It is much more difficult to design tubes that function both at low and high peak powers;
- (iv) All high power elements of the RFPC are passive, resulting in lower capital, maintenance and replacement costs per watt of peak power.

Let  $P_k$  be the peak input power to the PC network,  $T_k$  the time duration of that pulse,  $T_{cp}$  the duration of the compressed pulse,  $V_{ip}$  the section voltage without PC,  $V_{cp}$  the section voltage with PC, and  $T_f$  the section fill time. Define the beam energy multiplication factor  $M$ , the power multiplication factor  $P_g$ , the compression ratio  $C_r$  and the compression efficiency  $\eta_{pc}$  as

$$M = \frac{V_{cp}}{V_{ip}}; P_g = M^2; C_r = \frac{T_k}{T_{cp}}; \eta_{pc} = \frac{P_g}{C_r} \quad (1)$$

For a perfect compressor,  $P_g = C_r$  and  $\eta_{pc} = 1$ .

### SLED

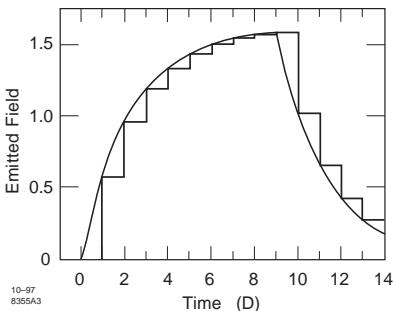


Figure 1: Time dependence of cavity field.

The functioning of SLED is best explained by considering the emitted field  $E_e$  and the reflected (output) field  $E_r$  of a resonant rf cavit. The basic SLED differential equation relates  $E_e$  to the input field  $E_i$ ,

$$\frac{dE_e}{dt} + qE_e = \frac{\alpha E_i}{T_L}; \quad q = \frac{1 - i \tan \psi}{T_L} \quad (2)$$

When the cavity is tuned,  $\psi = 0$ . Given the unloaded  $Q$ ,  $Q_0$ , and the external  $Q$ ,  $Q_e$ , of the cavities, then

$$\beta = \frac{Q_0}{Q_e}; Q_L = \frac{Q_0}{1 + \beta}; \alpha = \frac{2\beta}{1 + \beta}; T_L = \frac{Q_L}{\pi f}$$

$$\text{where } \tan \psi \equiv 2Q_L \frac{\Delta f}{f} \quad (3)$$

$\beta$  is the coupling coefficient,  $f$  is the resonant frequency of the SLED cavities.  $\alpha E_i \cos \psi$  is the steady state emitted field amplitude,  $\psi$  is the phase of the steady state emitted field.  $\alpha$  has a value between 0 and 2. For a constant  $E_i = a_0 e^{i\phi_0}$  and for an initial cavity emitted field  $E_e(t_0)$ , the solution of Eq.(2) is

$$E_e = E_{ef} + [E_e(t_0) - E_{ef}] e^{-q(t-t_0)} \quad (4)$$

$$E_{ef} = \alpha a_0 \cos \psi e^{i(\phi_0+\psi)}$$

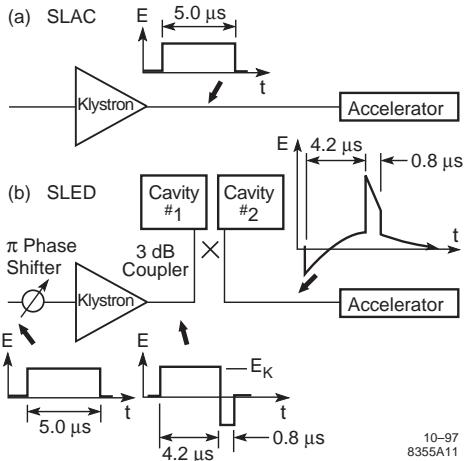


Figure 2: SLAC rf system (a) without SLED, (b) with SLED.

A typical emitted field is shown in Fig.1. The cavity output field  $E_r = E_e - E_i$ .

The SLAC rf system is shown in Fig.2, (a) without SLED and (b) with SLED. The 3-dB hybrid guides the output field  $E_r$  from both cavities into the accelerator section. During the first portion of the pulse from zero to  $T_1 = T_k - T_{cp}$  some energy is reflected but most of it flows into the cavities. At  $T_1$  we flip the input phase to the klystron by  $180^\circ$  causing the rf energy from the cavities and from the klystron to flow into the accelerator section. At the end of the rf pulse the cavities are nearly empty.

For an ideal SLED system, the cavities perfectly tuned, i.e.  $\psi = 0$ , the phase of the input to the PC network is shifted by  $\pi$  in zero time, and the compressed pulse duration equals the section fill time,  $T_{cp} = T_f$ . In that case, the normalized cavity output field during the compressed pulse is

$$E_r(t) = \sqrt{P(t)/P_k} = \gamma e^{-t/T_L} + E_{rss} \quad (5)$$

where

$$\gamma = \alpha(2 - e^{\tau_1}); \quad \tau_1 = \frac{(T_k - T_f)}{T_L}; \quad E_{rss} = 1 - \alpha$$

The SLED gain  $S_g$  defined as the unloaded beam energy with SLED divided by the unloaded beam energy without SLED, varies with time. Its maximum value  $S_{gm}$  is the integral of the normalized accelerating electric field along the section,  $E_r(t, z')$  at the end of the compressed pulse [1]

when  $t = T_f$ ,

$$M = S_{gm} = \int_0^t \left[ \gamma e^{(T_f - t(z'))/T_L} + E_{rss} \right] dz' \quad (6)$$

$$\text{where } z' = \frac{z}{L}; t(z') = T_f \frac{\ln(1+gz')}{\ln h}$$

$z$  is the distance along the section,  $L$  is the section length and the group velocity varies as  $v_g = v_{g0}(1+gz')$ . The ratio of the group velocity at the end of the section to the group velocity at the beginning is  $h = 1+g$ .

$$S_{gm} = \frac{\gamma e^{\tau_c(hx_{11}-1)}}{x_{11}(h-1)} + E_{rss} \quad (7)$$

$$\tau_c = \frac{T_f}{T_L}; x_{11} = 1 + \frac{\tau_c}{2\tau}$$

For a constant gradient section, such as the SLAC sections,  $h = e^{-2\tau}$ .

Choose  $Q_0$  such that  $Q_0/\pi f \gg T_k$ , as large as practicable and then choose  $Q_e$  to maximize  $S_{gm}$ . For SLED at SLC,  $T_k = 3.5 T_{cp}$ ,  $T_f = 0.82 \mu s$ ,  $\tau = 0.57$ . The SLED cavities are copper and operate in the  $TE_{015}$  mode. At the SLAC frequency of 2856 MHz, their theoretical  $Q_0$  is 108000. Their measured  $Q_0$  was about 100000. As a result of optimization and of practical considerations, the SLED parameters at SLAC are  $Q_0 = 100000$ ,  $Q_e = 20000$ . Thus  $\beta = 5$ ,  $\alpha = 1.667$ ,  $T_L = 1.86 \mu s$ . Assuming ideal SLED inputs:  $M = 1.614$  and  $\eta_{pc} = 0.61$ . The normalized SLED energy gain for a  $2.7 \mu s$  rf input pulse is shown in Fig.3.

We can obtain a nearly flat loaded beam energy by choosing an optimum current and modulating the input pulse for beam pulse lengths up to half a section fill time [4].

**SLED-II** In SLED-II, the two cavities are replaced by two resonant delay lines. Because the speed of EM waves is finite, the emitted field changes in discrete steps, with widths equal to the time it takes for a wave to transit down and back the length of the resonant line, resulting in a staircase emitted field waveform as shown in Fig.1. Most importantly, this results in a flat top output pulse. Because of this, and because the use of long low loss  $TE_{01}$  mode lines reduces the losses, and because its reflections during charging are less, SLED-II is more efficient than SLED. But, SLED-II is impractical at low frequencies because the length and diameter of the required resonant delay line are large. It becomes practical at X-band frequencies. A SLED-II system is considered for the NLC at 11.424 GHz.

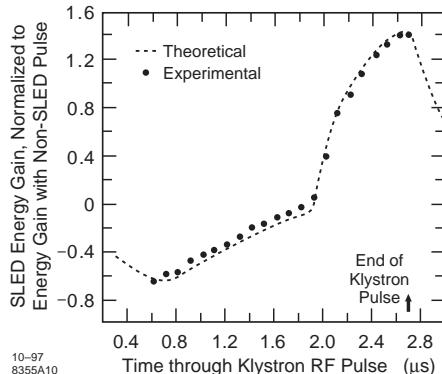


Figure 3: Normalized SLED energy gain.

A resonant delay line is a transmission line terminated in a short circuit and connected to an input transmission line via a coupling iris. The distance between the coupling iris and the short must be an integral multiple of half guide wavelengths, and the round-trip time delay  $T_d$  between the coupling network (iris) and the short equals the desired output pulse length. After turning on an incident field of amplitude  $E_i$ , the emitted field during the first round trip time is  $E_e(0) = 0$ . After  $nT_d$  time intervals it is given by [2]

$$E_e(n) = E_i \frac{(1-s^2)e^{-2\tau}}{1-se^{-2\tau}} (1-s^n e^{-2n\tau}) \quad (8)$$

$$= E_{ef}(1-s^n e^{-2n\tau}), \quad n = 1, 2, 3, \dots$$

Here,  $s$  is the reflection coefficient of the coupling network when the delay line is terminated in a matched load and  $\tau$  is the attenuation of the delay line in nepers. Superposition with the iris reflection yields

$$E_r(n) = E_e(n) - sE_i \quad (9)$$

The compression ratio  $C_r$  is an integer. At time  $t = (C_r - 1)T_{cp}$ , the phase of the input pulse is shifted by  $\pi$  so that the waves add and we get a compressed output of duration  $T_{cp} = T_b$  with the amplitude

$$E_{cp} = E_{ef} \left[ 1 - (se^{-2\tau})^{C_r-1} + s \right] E_i \quad (10)$$

For a given  $C_r$  and  $\tau$ ,  $s$  can be chosen to maximize  $E_{cp}$ . Examples of optimized  $s$ ,  $M$ , and  $\eta_{pc}$  values is shown in Tab.1.

The emitted field, output field and output power for  $C_r = 5$  are shown in Fig.4. A description of an optically switched RDL is given in [5].

## Sec.6.7: RF PULSE COMPRESSION

Table 1: Gain and efficiency examples as a function of compression ratio and RDL attenuation with SLED values for comparison.

| $C_r$ | $s$   | SLED $Q_0 = \infty$ |               | $\tau = 0.01$ |               |
|-------|-------|---------------------|---------------|---------------|---------------|
|       |       | $P_g$               | $\eta_{pc}\%$ | $P_g$         | $\eta_{pc}\%$ |
| 3     | 0.549 | 2.11                | 70            | 2.57          | 86            |
| 4     | 0.610 | 2.80                | 70            | 3.29          | 82            |
| 5     | 0.651 | 3.26                | 65            | 3.81          | 76            |
| 6     | 0.685 | 3.57                | 59            | 4.21          | 70            |

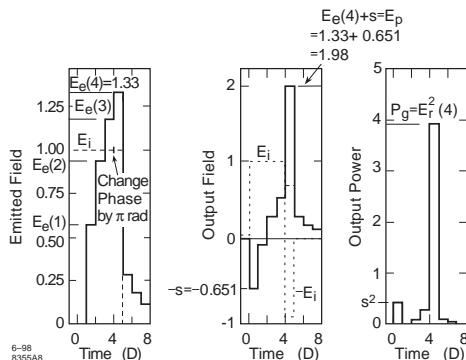


Figure 4: Emitted field, output field and output power.

**Binary pulse compression** As with SLED and SLED-II, the BPC modulation takes place at the low level input drive to the klystron amplifier, while the compression itself takes place at high power levels. But, unlike SLED and SLED-II, there are no reflection losses during charging. As with SLED-II, the dissipation loss is minimized by using as energy storing elements overmoded TE<sub>01</sub> circular guides. In this mode losses decrease as the 3/2 power of frequency for a fixed guide diameter and as the cube of the diameter for a fixed frequency. For example, at 11.4 GHz, a guide with an inside diameter of 2.81 in. has a loss of 1.1 dB per  $\mu\text{s}$  of time delay.

The basic component of the BPC is a 4 port 3-dB coupler with one of its ports connected to a delay line. Two pulses, each twice as wide as the delay line fill time, are applied to the two isolated ports of the coupler. During the first half of the input pulse, the relative phase of the inputs is such that all power exits the port connected to the delay line and is therefore delayed by half the input pulse width. At the beginning of the second half of the input pulse, the relative phase of the inputs

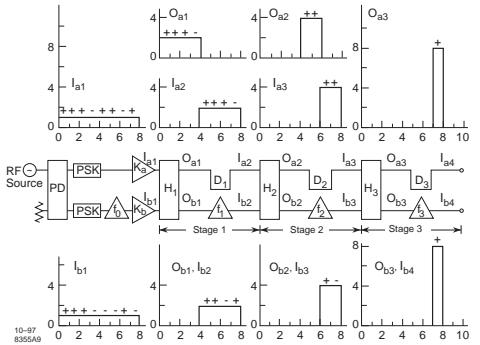


Figure 5: 3 stage BPC.

is reversed, causing all the power to exit the other port. Thus we have two output pulses, each half as wide and twice the power of each input pulse.

BPC stages can be cascaded. Thus, generally, a BPC consists of a series of stages, each stage consists of a 4-port 3 dB hybrid with one port followed by a delay line. The last delay equals the duration of the compressed pulse. Each additional delay toward the BPC input is double the previous delay. The two input pulses to the BPC are divided into  $2n$  equal bins. The length of each bin is the length of the compressed pulse. Each bin is coded with either a  $0^\circ$  or  $180^\circ$  phase shift. Consequently the output pulse length is divided by  $2n$  and the power is multiplied by  $P_g = 2n$  where  $n$  is the number of stages. The total required time delay per klystron is  $(C_r - 1) T_d / 2$ . A 3-stage BPC is shown in Fig.5. The BPC is more complex and requires longer delay lines than SLED-II. Both are considered for the NLC.

To reduce the delay line length, a variation of BPC, a multi-moded Delay Line Distribution System (DLDS) was proposed. Principle of DLDS is described in [6].

## References

- [1] Z.D. Farkas et al, Proc. 9th Int. Conf. High Energy Acc. (1976) p.576
- [2] P.B. Wilson et al, SLAC-PUB-5330 (1990)
- [3] Z.D. Farkas, IEEE Tran. MTT-34 (1986) p.1036
- [4] Z.D. Farkas, SLAC-AAS-note-83 (1994)
- [5] S.G.Tantawi et al, IEEE Trans. MTT-45 (1997) 1486
- [6] International Study Group Progress report, SLAC-R-559, KEK Report 2000-7 (2000)

## 6.7.2 Active Pulse Compression

*S. Tantawi, SLAC*

The SLED-II pulse compression system employs high  $Q$  resonant delay lines to store the energy during most of the duration of the incoming pulse. The round trip time of an rf signal through one of the delay lines determines the length of the compressed pulse. To discharge the lines, the phase of the incoming pulse is reversed  $180^\circ$  so that the reflected signal from the inputs of the lines and the emitted field from the lines add constructively thus, forming the compressed, high power, pulse.

The SLED-II system suffers from two types of losses that reduce its intrinsic efficiency. During the charging phase some of the energy is reflected at the delay line entrance, and never gets into the lines. Also, after the phase is reversed the energy inside the lines is not discharged completely in one compressed pulse time period. These two effects make the intrinsic efficiency of SLED-II deteriorate very fast at large compression ratios [1]. To reduce the amount of energy left-over after the output pulse is finished one can increase the coupling of the line just before the start of the output pulse. This will allow more energy to get out of the storage line during the compressed pulse time period. To reduce the losses due to reflections during the charging of the delay lines, one can optimize the constant line coupling for maximum energy storage.

The theory of active pulse compression with several time events is detailed in [2]. Here, we describe the special case of a *single event* switched pulse compression system.

### Passive pulse compression (Sec.6.7.1)



Figure 1: Resonant delay line.

Consider the waveguide delay line with a coupling iris shown in Fig.1. The *lossless* scattering matrix representing the iris is unitary. At a certain reference plane the matrix takes the following form

$$S = \begin{bmatrix} -R_0 & -j(1-R_0^2)^{1/2} \\ -j(1-R_0^2)^{1/2} & -R_0 \end{bmatrix} \quad (1)$$

In writing Eq.(1) we assumed a symmetrical structure for the iris two port network. The

forward and reflected fields around the iris are related,

$$\begin{aligned} V_1^- &= -R_0 V_1^+ - j(1-R_0^2)^{1/2} V_2^+ \\ V_2^- &= -j(1-R_0^2)^{1/2} V_1^+ - R_0 V_2^+ \end{aligned} \quad (2)$$

With the exception of some phase change, the incoming signal  $V_2^+$  at time  $t$  is the same as the outgoing signal  $V_2^-$  at time  $t-\tau$ , where  $\tau$  is the round trip delay through the line, i.e.

$$V_2^+(t) = V_2^-(t-\tau) e^{-j2\beta\ell} \quad (3)$$

where  $\beta$  is the wave propagation constant within the delay line, and  $\ell$  is the length of the line. Substituting Eq.(3) into Eq.(2) we get a difference equation that governs the system. During the charging phase we assume a constant input  $V_{in}$ . If the delay line has small losses ( $\beta$  has a small imaginary part), at resonance the term  $e^{-j2\beta\ell} = -p$  where  $p$  is a positive real number close to 1. We also assume that all the voltages are equal to zero at time  $t < 0$ . After the energy has been stored in the line, one may dump part of the energy in a time interval  $\tau$  by flipping the phase of the incoming signal just after a time interval  $(n-1)\tau$ . The output pulse level during the time interval  $(n-1)\tau \leq t < n\tau$  can be calculated, the result is

$$\begin{aligned} V_{out} &= V_1^-(n-1) \\ &= V_{in} \left[ R_0 + (1-R_0^2) \frac{1-(R_0 p)^{n-1}}{1-R_0 p} p \right] \end{aligned} \quad (4)$$

Indeed, this is the essence of the SLED-II pulse compression system. The maximum power gain of SLED-II is limited. Using Eq.(4), the maximum power gain as  $n \rightarrow \infty$  is,

$$\text{Max. power gain} = \frac{17}{p^2} - 8 - 12 \frac{\sqrt{2(1-p^2)}}{p^2} \quad (5)$$

which occurs at  $R_0 = \frac{1}{p} - \frac{\sqrt{8(1-p^2)}}{4p}$ . Clearly the maximum power gain is limited to 9 as  $p \rightarrow 1$ . Furthermore, this maximum is greatly affected by the losses in the delay line; for example, the gain is limited to 7.46 if the line has a 1% round trip power losses.

**Active switching during charging time** During the charging period the power reflected from the line reaches a maximum during the first time interval  $\tau$ . Hence, one could initially make the iris reflection coefficient zero. After the first time interval  $\tau$  we could switch the iris so that the reflection coefficient has a value  $R_0$ . Assuming a

resonant line and flipping the phase during the last time bin, the output pulse expression takes the form

$$V_{\text{out}} = V_{\text{in}} \left[ \frac{1 - (R_0 p)^{n-2}}{1 - R_0 p} (1 - R_0^2) p + (1 - R_0^2)^{1/2} p (R_0 p)^{n-2} + R_0 \right] \quad (6)$$

Again the choice of the value of  $R_0$  is such that  $V_{\text{out}}$  is maximized.

### Active switching during delay line discharge

#### Case 1: Discharging after the last time bin

To discharge the line, one can keep the input signal at a constant level during the time interval  $0 \leq t < n\tau$  but switching the iris reflection coefficient to zero so that all the energy stored in the line is dumped out. In this case,

$$V_{\text{out}} = V_{\text{in}} \frac{1 - (R_0 p)^n}{1 - R_0 p} (1 - R_0^2)^{1/2} p \quad (7)$$

#### Case 2: Switching just before the last time bin

To reduce the burden on the switch one can reverse the phase together with changing the iris reflection coefficient. In this case all the energy can still be dumped out of the line, but the iris reflection coefficient need not be reduced completely to zero. This new reflection coefficient is greater than zero and the switch need only change the iris between  $R_0$  and  $R_d$ . in this case the output reduces to

$$V_{\text{out}} = R_d V_{\text{in}} \left[ 1 + \left( \frac{1 - (R_0 p)^{n-1}}{1 - R_0 p} \right)^2 (1 - R_0^2) p^2 \right] \quad (8)$$

The required  $R_d$  to discharge the line completely is given by

$$R_d = \cos \left[ \tan^{-1} \left( \frac{1 - (R_0 p)^{n-1}}{1 - R_0 p} (1 - R_0^2)^{1/2} p \right) \right] \quad (9)$$

The compressed pulse takes place in the interval  $(n-1)\tau \leq t < n\tau$ . The optimum value of  $R_0$  is such that it fills the system with maximum possible amount of energy in the time interval  $(n-1)\tau$ . Fig.2 shows an example of this case in comparison with a normal sled pulse compression system.

Unlike the passive system, the maximum power gain has no intrinsic limit. It is only limited by the amount of losses in the storage line. In this case the gain can be much higher than 9, which is the limit of the passive system.

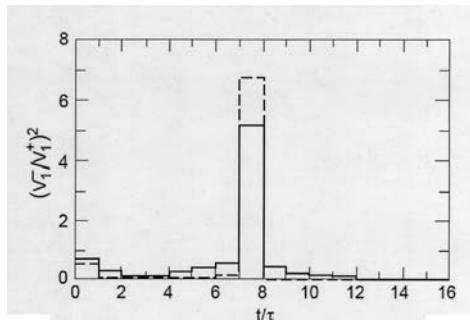


Figure 2: Comparison between SLED-II output (solid curve) and a one time switched resonant delay line (dashed curve) for a compression ratio of 8. The line is switched just before the last time bin.

**Comparison** Table 1 compares the different types of pulse compression systems. It also gives the optimum system parameters for each compression ratio  $C_r$ ; here  $C_r$  is defined as the total time interval divided by the duration of the compressed pulse, i.e.  $n\tau$ . The efficiency of the system  $\eta$  is defined as the energy in the compressed pulse divided by the total incident energy. In these calculations we assume a lossless system, i.e.  $p = 1$ .

## References

- [1] P.B. Wilson, Z.D. Farkas, R.D. Ruth, Linear Accl. Conf. 1990, SLAC-PUB-5330
- [2] S.G. Tantawi et al, NIM A370 (1996) 297

### 6.7.3 Ultra-High-Power Multimoded Rf Components

S. Tantawi, C. Nantista, SLAC

The high power requirements of the rf systems for linear colliders stimulated significant development in passive microwave components [1, 2]. The power handling capabilities of such components have increased considerably [2]. This has been achieved by increasing the size of the components with respect to the operating wavelength, i.e. by overmoding these components. In particular, a class of microwave structures with complete planar symmetry has been developed. These components carry only  $\text{TE}_{n0}$  modes. This makes it possible to make all the manipulations in the 2-D plane. The height of these components can

Table 1: Comparison between different methods of single event switching pulse compression systems.

| $C_r$ | SLED-II    |            | Switching during charging Time |            | Discharging after the last time bin |            | Discharging just before the last time bin |            |       |
|-------|------------|------------|--------------------------------|------------|-------------------------------------|------------|---|------------|-------|
|       | $\eta(\%)$ | Opt. $R_0$ | $\eta(\%)$                     | Opt. $R_0$ | $\eta(\%)$                          | Opt. $R_0$ | $\eta(\%)$                                | Opt. $R_0$ | $R_d$ |
| 2     | 78.1       | 0.5        | 100                            | 0.707      | 84.4                                | 0.5        | 100                                       | 0.0        | 0.707 |
| 3     | 88.7       | 0.548      | 98.9                           | 0.631      | 82.7                                | 0.646      | 89.6                                      | 0.5        | 0.610 |
| 4     | 86.0       | 0.607      | 92.6                           | 0.658      | 82.1                                | 0.725      | 87.0                                      | 0.646      | 0.536 |
| 5     | 80.4       | 0.651      | 85.1                           | 0.688      | 81.9                                | 0.775      | 85.7                                      | 0.725      | 0.483 |
| 6     | 74.6       | 0.685      | 78.1                           | 0.714      | 81.8                                | 0.809      | 84.9                                      | 0.775      | 0.443 |
| 8     | 64.4       | 0.733      | 66.5                           | 0.754      | 81.6                                | 0.854      | 84.0                                      | 0.835      | 0.386 |
| 10    | 56.2       | 0.767      | 57.7                           | 0.783      | 81.6                                | 0.882      | 83.4                                      | 0.869      | 0.346 |
| 12    | 49.9       | 0.792      | 50.9                           | 0.805      | 81.5                                | 0.900      | 83.1                                      | 0.892      | 0.317 |
| 16    | 40.6       | 0.828      | 41.2                           | 0.837      | 81.5                                | 0.924      | 82.7                                      | 0.920      | 0.275 |
| 24    | 29.6       | 0.869      | 29.8                           | 0.875      | 81.5                                | 0.949      | 82.2                                      | 0.947      | 0.225 |
| 32    | 23.3       | 0.893      | 23.4                           | 0.897      | 81.5                                | 0.961      | 82.0                                      | 0.960      | 0.195 |
| 64    | 12.6       | 0.936      | 12.7                           | 0.938      | 81.5                                | 0.981      | 81.7                                      | 0.980      | 0.138 |
| 128   | 6.6        | 0.962      | 6.6                            | 0.963      | 81.5                                | 0.990      | 81.6                                      | 0.990      | 0.099 |
| 256   | 3.4        | 0.978      | 3.4                            | 0.979      | 81.5                                | 0.995      | 81.5                                      | 0.995      | 0.069 |

then be increased to reduce the field and accommodate high power operation. This class of components is overmoded in both height and width, allowing simultaneous manipulation of multiple  $TE_{n0}$  modes; i.e. multimoding. Such components can thus be made to perform multiple functions, resulting in compact and efficient system integration. To connect between these components and the circular waveguides needed for efficient transfer of rf power over long distances, a special type of multimoded circular-to-rectangular taper was developed.

Another class of multimoded components that depends on the azimuthal symmetry of circular waveguide carrying  $TE_{0n}$  modes has been developed. This class is used to reduce the length of rf storage lines. The designs and functions of some planar rf components are presented below. For details on overmoded circular components, see [1].

**Planar components** The initial idea of using overmoded planar components was motivated by the need for a 3dB hybrid capable of handling hundreds of megawatts of X-band rf power in a SLED-II pulse compression system [3]. A satisfactory device was designed based on the R.H. Dicke's circuit synthesis of a 3dB hybrid [4]. It was soon realized that the two interconnecting guides could be combined into one single guide carrying two modes [5], as shown in Fig.1. This

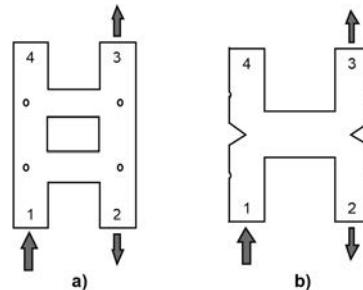


Figure 1: Schematic of the H-planar geometries of the a) two-rung ladder and b) "magic H" hybrid designs. Power-flow arrows indicate output ports for the indicated input port.

resulted in the so-called magic-H hybrid. The device has full planar symmetry, and the height can be adjusted to any value required to reduce surface field and increase power handling capacity. Magic-H hybrids have been used in high power rf systems.

If two hybrids of this type are merged, side-by-side, leaving the miters to form a post, the combined ports can produce dual moded waveguides with the same width as the connecting guide (Fig.2). Adding properly spaced T-junctions to the new ports of this device resulted in the invention of the so-called cross potent superhybrid [6], illustrated in Fig.2.

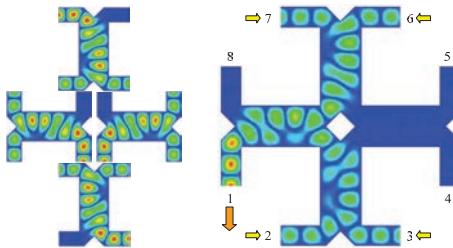


Figure 2: Combining four magic-H hybrids resulted in the invention of the Cross Potent Superhybrid.

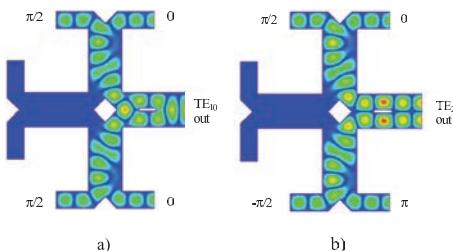


Figure 3: Cross Potent Launcher with simulated electric field plots illustrating launching a)  $\text{TE}_{10}$  and b)  $\text{TE}_{20}$  in the right overmoded rectangular port with the indicated relative phases for four equal amplitude inputs. Alternate phasings of the inputs send the power to either of the left ports.

As indicated in Fig.2, this superhybrid can be used to combine power from four rf sources into any one of four outputs. The choice of the output port is determined by the phases of the inputs. The scattering matrix of the device is

$$S = \frac{1}{2} \begin{bmatrix} 0 & 1 & -i & 0 & 0 & -1 & -i & 0 \\ 1 & 0 & 0 & -i & -1 & 0 & 0 & -i \\ -i & 0 & 0 & 1 & -i & 0 & 0 & -1 \\ 0 & -i & 1 & 0 & 0 & -i & -1 & 0 \\ 0 & -1 & -i & 0 & 0 & 1 & -i & 0 \\ -1 & 0 & 0 & -i & 1 & 0 & 0 & -i \\ -i & 0 & 0 & -1 & -i & 0 & 0 & 1 \\ 0 & -i & -1 & 0 & 0 & -i & 1 & 0 \end{bmatrix} \quad (1)$$

One variation on this device is achieved by eliminating two of the output ports in exchange for a single port carrying two modes. Fig.3 shows such a device, capable of launching either the  $\text{TE}_{10}$  or the  $\text{TE}_{20}$  mode into the overmoded port, depending on the phases of the input devices. Dealing with two modes in the same system is thus possible. For example, Fig.4 shows an H-plane bend that transfers both of these modes perfectly. For the theory for this bend, see [7, 8].

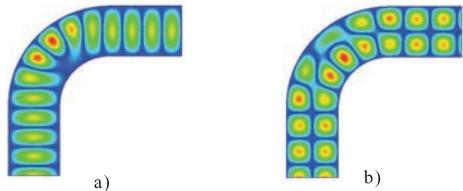


Figure 4: Overmoded H-plane bend waveguide with simulated electric field plots illustrating a)  $\text{TE}_{10}$  mode transmission and b)  $\text{TE}_{20}$  mode transmission.

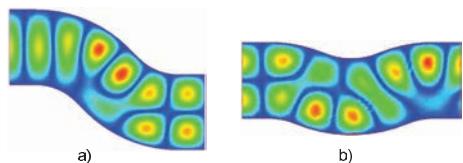


Figure 5: a) Jog mode converter and b) 50/50 mode mixer with field patterns.

Conversion between these modes or mixing can be accomplished with a simple jog, bend, or wiggle, as illustrated in Fig.5. By adding mixers to overmoded ports on the central junction of the cross potent superhybrid, one can make a device that separates these modes into different waveguides. Fig.6 shows this device and an alternate mode separator design which uses a  $45^\circ$  bend for mixing. Such a device allows power to be sent through the same transmission waveguide to different destinations via different modes.

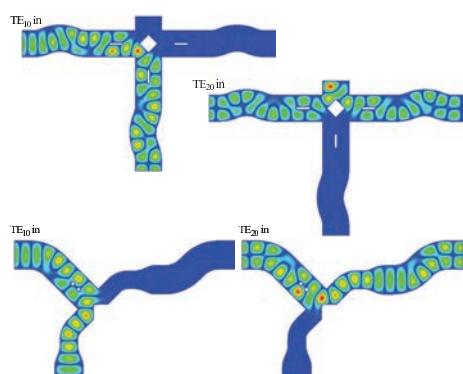


Figure 6: Use of mode mixers and mode converters in mode separators/extractors.

**Circular-to-rectangular transition** Integrating planar overmoded rectangular waveguide components into systems using overmoded circular waveguide for low-loss power transmission and energy storage requires a mode converter. A special type of compact circular-to-rectangular taper [9] has been designed for this purpose. This taper converts the rectangular guide  $TE_{10}$  mode into the fundamental  $TE_{11}$  mode of circular guide and the rectangular  $TE_{20}$  mode into the low-loss  $TE_{01}$  mode of the circular guide (see Fig.7). Details on the design of this taper are described below.

Smooth taper We require that taper sections be buildable using wire Electrical Discharge Machining (EDM). When tapering from one shape, e.g. a circle, to another shape, e.g. a rectangle, the length of the taper  $\ell$  and the connecting points between the two shapes uniquely define the taper. In cylindrical coordinates a shape  $i$  placed with cylindrical symmetry around the  $z$ -axis can be described by a relation  $r_i(\phi)$ , which gives the radius as a function of the angle  $\phi$ . The taper between two shapes  $r_1(\phi)$  and  $r_2(\phi)$  is then given by

$$r(\phi, z) = r_1(\phi) + \frac{r_2(\phi) - r_1(\phi)}{\ell} z \quad (2)$$

This taper is compatible with the process of wire EDM when the two heads of the machine are moving synchronously with the same angular speed. More complicated tapers are described by a set of tapers, each of the form of Eq.(2), cascaded together.

Modal connectivity and waveguides cross-section dimensions First, let us consider an adiabatic taper between a square waveguide and

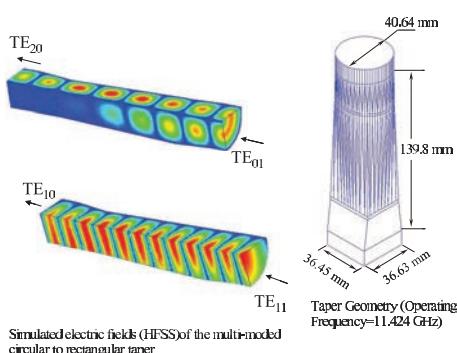


Figure 7: Dual-moded circular-to-rectangular mode converter/taper.

Table 1: Mode mapping between the first few modes for a sufficiently adiabatic taper between a circular waveguide and a square waveguide.

| Circular Guide modes       | Square Guide Modes                            |
|----------------------------|---|
| $TE_{11}$ (Polarization 1) | $TE_{10}$                                     |
| $TE_{11}$ (Polarization 2) | $TE_{01}$                                     |
| $TM_{01}$                  | $TM_{11}$                                     |
| $TE_{21}$ (Polarization 1) | $TE_{20}$ and<br>$TE_{02}$ ( $\pi$ out Phase) |
| $TE_{21}$ (Polarization 2) | $TE_{11}$                                     |
| $TE_{01}$                  | $TE_{20}$ and<br>$TE_{02}$ (in phase)         |
| $TM_{11}$ (Polarization 1) | $TM_{12}$                                     |
| $TM_{11}$ (Polarization 2) | $TM_{21}$                                     |
| $TE_{31}$ (Polarization 1) | $TE_{12}$                                     |
| $TE_{31}$ (Polarization 2) | $TE_{21}$                                     |
| $TM_{21}$ (Polarization 1) | $TM_{22}$                                     |
| $TM_{21}$ (Polarization 2) | $TM_{31}$ and<br>$TM_{13}$ ( $\pi$ out phase) |
| $TE_{41}$ (Polarization 1) | $TE_{22}$                                     |
| $TE_{41}$ (Polarization 2) | $TE_{31}$ and<br>$TE_{13}$ (in phase)         |
| $TE_{12}$ (Polarization 1) | $TE_{30}$                                     |
| $TE_{12}$ (Polarization 1) | $TE_{03}$                                     |
| $TM_{02}$                  | $TM_{31}$ and<br>$TM_{13}$ (in phase)         |

a circular waveguide. Adiabaticity can be insured by increasing the length  $\ell$  in Eq.(2) as long as necessary. With the dimensions of the square and circular waveguides chosen to be as close as possible, so that they support approximately the same number of modes, the taper will map the modes between the two waveguides according to Tab.1. The taper simply preserves the common reflection symmetries. The  $S$ -matrix of the transition connects modes of the same symmetry class, and for a sufficiently adiabatic transition preserves their TE or TM character. The most likely modes of interest are the  $TE_{11}$  and the  $TE_{01}$  in the circular waveguide. These map, respectively, to the  $TE_{10}$  (or  $TE_{01}$ ) and a combination of the  $TE_{20}$  and  $TE_{02}$  in the square waveguide.

Dual-moded taper Adiabaticity is sufficient to map the  $TE_{11}$  in circular guide to  $TE_{10}$  in the square waveguide. Indeed, this is well known. However, one would want to convert the  $TE_{01}$  mode in the circular guide to a single polarization of  $TE_{02}/TE_{20}$  in the square guide. Modifying the

square waveguide to a rectangular waveguide to break the degeneracy between the TE<sub>02</sub> and TE<sub>20</sub> modes could do this. However, in this case, the length of the taper required to achieve an adiabatic transition to a single mode in the rectangular guide is excessive. If one chooses the dimensions of the square waveguide and the circular waveguide such that the TE<sub>02</sub> circular mode and the TE<sub>22</sub> and TM<sub>22</sub> rectangular modes do not propagate, the length of this taper can be approximately 18 cm at 11.4 GHz; i.e. about 7 wavelengths.

Instead, a taper for NLC development was constructed from three sections. The middle section is a cylinder with the following shape

$$r_2(\phi) = r_0(1 + 0.1 \cos 2\phi) \quad (3)$$

where  $r_0$  is the radius of the circular guide. The dimensions of the rectangular guide are chosen such that the rectangular cross section satisfies

$$r_3(\phi) = r_0(1 + 0.1 \cos 2\phi + \sum_{i=3}^{\infty} \alpha_i \cos i\phi) \quad (4)$$

where the coefficients  $\alpha_i$ 's are chosen to produce a rectangular shape.

The taper from the circle to the intermediate shape scatters the TE<sub>01</sub> mode into two modes M<sub>1</sub> and M<sub>2</sub> in the intermediate section. Also the taper from the rectangular waveguide to the intermediate shape scatters the rectangular mode TE<sub>02</sub> into the same M<sub>1</sub> and M<sub>2</sub> modes. The lengths of both tapers are adjusted such that the coefficients of the scattered modes, M<sub>1</sub> and M<sub>2</sub>, are the same from both sides. Since M<sub>1</sub> and M<sub>2</sub> propagate with different phase velocities in the intermediate section, the length of that section could be adjusted so that the circular TE<sub>01</sub> mode gets completely converted into the rectangular TE<sub>02</sub> mode. The idea of this design was first reported in [8]. It was later implemented to split the TE<sub>01</sub> output of a gyrokylystron. However, in both cases, no care was taken to insure the adiabatic propagation of the TE<sub>11</sub> mode. Because of the odd shape in the middle of this taper, described by Eq.(3), the taper tended to couple the rectangular TE<sub>01</sub> to the circular TE<sub>31</sub> mode. This is not surprising, since that shape essentially contains a second order azimuthal deformation which tends to couple modes that differ in azimuthal index by 2. In the other direction, the circular TE<sub>11</sub> modes also coupled to the TE<sub>12</sub> mode in the rectangular guide.

The design shown above was arrived at after several design iterations, increasing the length until near perfect conversion between one polarization of the TE<sub>11</sub> circular mode and the TE<sub>01</sub> (not TE<sub>10</sub>) rectangular mode was achieved. Each time the length was increased, the full design was repeated to insure perfect conversion between the TE<sub>02</sub> rectangular mode and the TE<sub>01</sub> circular mode. Then the TE<sub>11</sub> circular to TE<sub>01</sub> rectangular transmission was checked.

For examples of further developments in high-power rf component designs for accelerator applications, see [10, 11, 12, 13, 14, 15].

## References

- [1] S.G. Tantawi et al, PRST-AB 5 (2002)
- [2] S.G. Tantawi et al, PRST-AB 8 (2005) 042002
- [3] P.B. Wilson, Z.D. Farkas, R.D. Ruth, Linear Accel. Conf. (1990)
- [4] S.G. Tantawi et al, IEEE Trans. Microwave Theory Tech., vol.47, no.12 (1999) 2539
- [5] C.D. Nantista et al, PAC 99
- [6] C.D. Nantista, S.G. Tantawi, IEEE Microwave Guided Wave Lett., vol.10, no.12 (2000) 520
- [7] C.D. Nantista, S.G. Tantawi, Proc. Advanced Accelerator Concepts (2000) p.702
- [8] S.G. Tantawi, J. Electromagnetic Waves & Applications vol.18, no.12 (2004) 1679
- [9] S.G. Tantawi, N.M. Kroll, K. Fant, PAC 99
- [10] I. Syratchev and W. Bruns, 2001 Particle Accelerator Conference (PAC'01), Chicago, Ill., U.S.A., June 18-22, 2001.
- [11] C.D. Nantista, 6th Workshop on High Energy-Density and High Power RF (RF 2003), Berkeley Springs, West Virginia, June 22-26, 2003; SLAC-PUB-10218.
- [12] M.Yedduula, S. Tantawi et. al, IEEE Transactions On Microwave Theory and Techniques, June 2009.
- [13] C.D. Nantista, C. Adolphsen, 2006 Linear Accelerator Conference (LINAC'06), Knoxville, TN, Aug. 21-25, 2006; SLAC-PUB-12372.
- [14] S. Kazakov et al, 11th European Particle Accelerator Conference (EPAC'08), Genoa, Italy, June 23-27, 2008. C. Nantista, C. Adolphsen, 2009 Particle Accelerator Conference (PAC'09), Vancouver, B.C., Canada. May 4-8, 2009; SLAC-PUB-13696.
- [15] I. Syratchev, A. Cappelletti, 1st International Particle Accelerator Conference (IPAC'10), Kyoto, Japan, May 23-28, 2010.

## 6.8 RF WINDOWS AND CAVITY COUPLING

R.M. Sundelin, H.L. Phillips, TJNAF

**Window types and characteristics** Fig.1 shows 8 common window types.

Their relative advantages (+) and disadvantages (-) are: (i) +compact, bias to stop multipactor, simple geometry, relatively broad band, - inner braze problematic, enhanced energy density, prone to multipacting, high voltage gradient, center conductor cooling; (ii) + reduced energy density, strong, broad band, bias to stop multipactor, - inner braze problematic, high voltage gradient, center conductor cooling, complex shape; (iii) + reduced energy density, - center conductor cooling, difficult to fabricate, local standing wave; (iv) + simple braze geometry, somewhat reduced energy density, - complex field patterns, local standing wave; (v) +moderately broad band, low power density, - enhanced energy density, corner brazes problematic, corners require rounding; (vi) + simple seal, can use compression seal, - enhanced energy density, local standing wave, narrow band; (vii) + broad band, - enhanced energy density, high local fields, corner brazes problematic, corners require rounding; (viii) + simple window geometry, - local space harmonics, high power density, enhanced energy density.

**Frequency scaling** If the transmission line and window dimensions are scaled as  $f^{-1}$  and the loss tangent is independent of  $f$  the power handling capabilities are affected as

| Limitation                  | cooling         | $f$ scaling          |
|-----------------------------|-----------------|----------------------|
| Field                       | -               | $f^{-2}$             |
| Peak temp.<br>(runaway)     | edge<br>surface | $f^{-1}$<br>$f^{-2}$ |
| Temp gradient<br>(cracking) | edge<br>surface | $f^{-2}$<br>$f^{-3}$ |

**Window cooling methods** Natural convection is the simplest and is often used but has limited effectiveness. Forced air improves the effectiveness and is relatively simple to employ but entrained dirt can be a problem. Surface cooling by a low loss dielectric fluid, such as Freon 113 is highly effective. (Note that Freon 113 is now no longer permitted for environmental reasons. Some of the fluids mentioned in Sec.7.2.7 may prove to be as effective.)

**Window matching methods** Iris matching is simple but leads to field concentrations with

overheating and multipacting potential.  $\lambda_g/2$  is effective but entails increased power absorption. Two windows at  $\lambda_g/4$  is effective, provides redundancy and surface coolant confinement but is complex.

**Failure modes** (Steps to avoid or prevent a particular mode are labeled PV while those giving protection from equipment damage in the event of the phenomenon are labeled PT.) *Thermal gradients* can lead to stress fracture. Good cooling, dense ceramic with no voids or binder clusters and elimination of multipacting PV. Infrared interlock, PT. *High temperature* with attendant loss tangent increase can lead to thermal runaway. Good cooling PV and infrared interlocks PT. *Multipacting* leads to unwanted heating of ceramic. Anti multipactor coatings, bias of center conductor, good design (see Sec.6.9) PV, electron detectors and infrared interlock PT. *Charging* is caused by impact on window of photons or charged particles which can initiate arcs leading to window puncture. Shielding of the window from beam and field emission sources together with a low loss resistive coating for charge drainage PV. Arc and vacuum interlocks PT. *Gas discharge* can lead to rf sputtering of metal onto the window with consequent heating and breakage. Avoidance of multipacting, and maintenance of  $< 10^{-7}$  Torr pressure PV. Arc, infrared, vacuum and VSWR interlocks PT. *Edge arcing* occurs at braze locations. Avoidance of braze recess or protrusion PV. Arc detector PT. *Ceramic breakdown* can puncture the ceramic and requires avoidance of ceramic defects PV. *Higher order mode effects* can be caused by resonances at the window which lead to heating and gas discharges. Avoidance of resonant grooves associated with corona shields PV.

**Design details** To avoid thermal stress cracking for a given ceramic shape and loss tangent, a sequence of simulations is used to calculate the EM fields, heat source distribution, temperature map, and stress field. A minimum deployment of ceramic material can be used in deriving the window shape in order to withstand the atmospheric load and the superimposed thermal stress. Surface heating from an antimultipacting coating must be included if Ti N or other conductive materials are used.

**Comparison of window materials** Factors affecting thermal stress failure are thermal conductivity,  $K$ , thermal expansion coefficient,  $C$ , loss

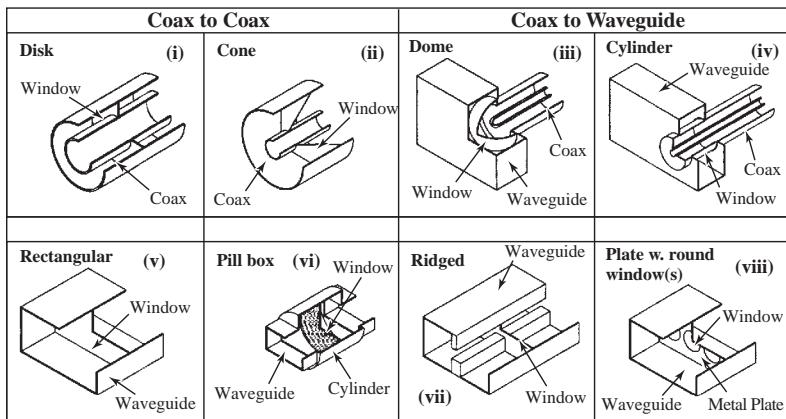
7-98  
8355A271

Figure 1: 8 common window types.

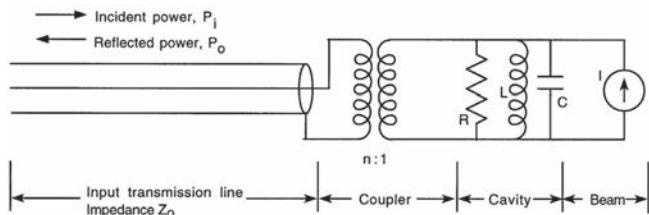
**Equivalent Circuit of Cavity, Input Coupling, and Beam**

Figure 2: Equivalent circuit.

tangent,  $\tan \delta$ , young's modulus,  $E$ , permittivity,  $\epsilon$ , and breaking stress,  $S_b$ . An approximate figure of merit for a given RF field value and frequency [1] can be written as:

$$m = \frac{K \cdot S_b}{C \cdot \tan \delta \cdot E \sqrt{\epsilon}}. \quad (1)$$

Rough values of  $m$  (normalized to 100) for a few materials are given in the table "Material Comparison" below. However, values of  $\tan \delta$  can vary considerably depending on vendor and process.

**Metallization Losses** The RF surface impedance of Molybdenum- Manganese metallization is not well quantified and varies substantially. New active alloy braze materials such as Ag (97%), Cu (1%), Zr (2%) (under development, Wesgo Metals) with substantially higher electrical conductivity and good wetting of ceramics are available [2], [3].

**Examples** In the table below, examples of achieved power coupling to klystron and accelerator cavities are given.

| Usage            | $f[\text{GHz}]$ | $P[\text{MW}]$ | Fig.1        |
|------------------|-----------------|----------------|--------------|
| Klystron, pulsed | 11.4            | 58 [4]         | (v),circular |
| Klystron CW      | 0.35            | 1.3 [5]        | (i)          |
| Cavity, pulsed   | 2.856           | 310 [6]        | (vi)         |
| Cavity, CW       | 0.5             | 1.0 [7]        | (iv)         |

**Fabrication details** Since the ceramic is brittle, any metal rigidly connected to it must have a low enough yield strength and/or be thin enough that stresses exerted on the ceramic due to differential thermal expansion (including those that occur during brazing) do not exceed the ceramic strength. Alloys, such as Kovar, (ASTM 1466 grade), with expansion coefficients similar to that of the ceramic are sometimes used as the metal which is brazed to the ceramic.

Metallization of the ceramic edges is normally accomplished using a commercial

## MATERIAL COMPARISON

| Material        | $\epsilon$ | $\delta$             | K    | C   | E   | Sb  | m      |
|-----------------|------------|----------------------|------|-----|-----|-----|--------|
| 99.5% $Al_2O_3$ | 9.58       | $9 \times 10^{-5}$   | 29.3 | 6.9 | 372 | 310 | (16.5) |
| 97% $Al_2O_3$   | 9.53       | $4.5 \times 10^{-4}$ | 26.8 | 6.9 | 350 | 296 | (15.3) |
| $Si_3N_4$       | 7.5        | $2 \times 10^{-3}$   | 27   | 2.3 | 503 | 665 | (82)   |
| AlN             | 8.5        | $3 \times 10^{-3}$   | 115  | 5.7 | 300 | 300 | (100)  |

molybdenum manganese mixture. For ceramics (See Sec.5.10.8) having a low percentage of glass phase, glass is included in the metallization to ensure adequate bonding to the ceramic.

Brazing of the metallized ceramic to the adjacent metal structure is done using a braze which is compatible with both the metallization and the metal structure. Copper and gold are examples of braze materials used.

**Cavity coupling** (See also Secs.7.3.6, 7.3.7, 7.3.7, 7.3.8, 7.3.8.1)

Common coupler types The common coupler types with their relative advantages (+) and disadvantages (-) are:

*Coaxial (magnetic loop)* + compact, adjustable (not under vacuum), rigid against vibration, - cooling shorted end while maintaining de-mountability difficult, high power density on center conductor;

*Coaxial (electric probe)* + compact, easily de-mountable, adjustable under vacuum, - subject to vibration, high power density on center conductor, coolant must be introduced and withdrawn through center conductor;

*Waveguide* + low power density, no center conductor, - large (can be reduced by ridging);

*On axis coupling* + no enhancement of cell wall current density, does not induce multipacting in cells, - uses additional beam line space, coupling limited;

*Through cell-wall coupling* + minimizes beam line space required, strong coupling, - enhances cell wall current density lowering Q, causing hot spots and perhaps inducing multipacting in cells.

Power-voltage -current relationships

(See also Secs.2.4.3, 6.18) See Fig.2. In the absence of beam loading, with a maximum energy gain in the cavity (length L)  $V_p$  [volts] and with a transit time adjusted shunt impedance  $R$  the power dissipated in the cavity wall is  $P = V^2/RL$  which can be taken as a definition of  $R$  ("linac" definition, i.e. no factor of 2).

For a high Q cavity in which the input coupler is adjusted to provide no reflection of the drive power when the average beam current is  $I_0$  at a peak cavity voltage  $V_{p0}$  the required incident drive power for other currents and voltages, assuming beam at voltage crest is

$$P_i = V_{p0} I_0 (V/V_{p0} + I/I_0)^2 / 4 \quad (2)$$

For a standing wave cavity which is tuned to a frequency that maximizes  $V_p$ , the  $V$ ,  $V_p$ , and  $P_r/P_i$  where  $P_r$  in the power reflected and  $P_i$  the incident power, are given by [8]

$$V_p = \sqrt{P_i RL} \frac{2\beta^{1/2}}{1+\beta} \left[ 1 - \frac{I}{2} \sqrt{\frac{RL\beta}{P_i}} \cos \phi \right] \quad (3)$$

where  $\phi$  is the bunch - crest phase difference.  $V = V_p \cos \phi$  and the power ratio is

$$\frac{P_r}{P_i} = 1 - \frac{IV}{P_i} - \left[ \frac{4\beta}{(1+\beta)^2} \right] \left[ 1 - \frac{I}{2} \sqrt{\frac{RL}{\beta P_i}} \cos \phi \right]^2 \quad (4)$$

If multiple probes have appreciable coupling to a cavity,

$$Q_L^{-1} = Q_0^{-1} + Q_{ext1}^{-1} + Q_{ext2}^{-1} + \dots Q_{ext-n}^{-1} \quad (5)$$

For VSWR relations see Secs.6.6, 6.16.

## References

- [1] L. Doolittle, Lawrence Berkley National Lab
- [2] Walker, V.C. Hodges, Welding Journal, AWS, Oct. 2008, P 43-50
- [3] American Welding Soc. Brazing Handbook, 5th edition 2007
- [4] P.B. Wilson, PAC 95, p.1487
- [5] E.-G. Schweiße, R. Bachmor, E. Demmel, PAC 93, p.1178
- [6] A. Miura, H. Matsumoto, PAC 93, p.1124
- [7] R. Sundelin et al PAC 81, p.2884
- [8] P.B. Wilson, C.S. Nunan, PAC 73, p.1021

## 6.9 MULTIPACTING

D. Proch, J. Sekutowicz, DESY

Multipacting describes a vacuum phenomenon of resonant electron multiplication: electron trajectories are determined by the electric and magnetic RF fields. When impacting the surface one electron might create more than one secondary, depending on the impact energy and material. If the time of flight is a multiple of the RF period, an electron avalanche will be initiated and sparking might result. In a super-conducting (sc) cavity the impact energy of multipacting current might initiate a local thermal instability (quench).

**Two point multipacting between parallel plates with only electric field acting** The resonant condition is given by

$$V_n = \frac{7.15}{2n-1} d^2 f^2 \quad (1)$$

where  $V_n$  [kV] is the peak voltage at frequency  $f$  [GHz] at which resonance occurs in a gap of  $d$  [cm] and  $n$  being the “order” i.e. how many RF periods between electron impacts.

**One point multipacting [1]** This type of multipacting occurs for example at the cylindrical wall of a  $\text{TM}_{010}$  resonator (Fig.1). The trajectories are determined by the bending force of the magnetic field component  $H_n$ . The electric field is small and does not influence the resonance conditions.

$$H_n \approx 220 \cdot \frac{f}{n} [\text{Oe}] \quad (2)$$

The impact energy of the electron at the surface is

$$W_{\text{imp}} \approx 7.6 \cdot 10^3 \cdot E_r^2 / f^2 [\text{eV}] \quad (3)$$

where  $E_r$  [MV/m] is the perpendicular electric field and  $f$  is as above the driving frequency in [GHz]

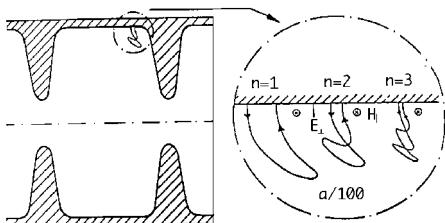


Figure 1: One point multipacting in cylindrical cavities ( $a$ : radius of cavity,  $n$ : order of multipacting).

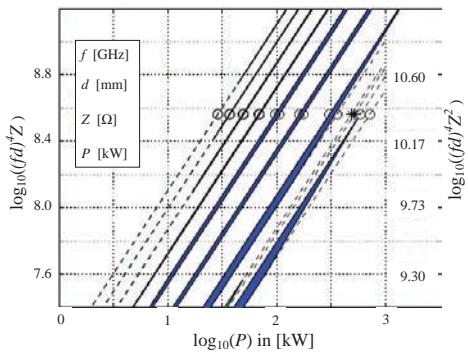


Figure 2: Multipacting in coaxial lines.

**Multipacting in a coaxial line [2]** is analyzed by tracking codes: electrons start at, and return to, the outer conductor without touching the inner conductor (although crossing nearby the coax gap). There are also two point trajectories between outer and inner conductor, but the impact energy does not create enough secondary electrons in most cases. A design chart for avoidance of multipacting in coaxial lines is shown Fig.2.

It is a graphical summary of the multipacting analysis in coaxial lines including the scaling laws and the secondary yield. The left vertical axis is for one-point multipacting, marked by circles in the logarithm of  $(fd)^4 Z$  (in  $(\text{GHz} \cdot \text{mm})^4 \cdot \Omega$ ), and the right vertical axis is for two-point processes, the processes marked by asterisks, in the logarithm of  $(fd)^4 Z^2$  (in  $(\text{GHz} \cdot \text{mm})^4 \cdot \Omega^2$ ). The horizontal axis is the logarithm of incident power (for all processes) in kW.  $Z$  is the impedance of coaxial line,  $d$  is the distance between inner and outer conductor and  $f$  is frequency of the propagating wave. The lowest band (at right) is the first order one-point multipacting band. The next band upwards is a two-point first order band, followed by a set of one-point multipacting bands. The order of multipacting increases up to eight when in the figure one moves towards lower power (to left). The circles (one-point) and asterisks (two-point) depict the case of  $50 \Omega$ ,  $40 \text{ mm}$  (outer diameter) coaxial line transmitting  $1.3 \text{ GHz}$  wave. Dark blue parts of the bands represent areas where multipacting may occur, i.e. the average impact energy is in the range  $\approx 100-1500 \text{ eV}$  and the secondary yield is  $> 1$ .

**Scaling laws for one point trajectories** For standing waves, when the reflection coefficient

$\rho = 1$ , multipacting power levels scale as follows:

$$P_{SW} \propto f^{frac}(d \cdot f)^4 Z(n+1)^2, \quad E_{kin} \propto f^2 d^2 \quad (4)$$

where  $P_{SW}$  is the standing wave power. When the reflection coefficient  $\rho < 1$ ,

$$P_\rho \approx \left[ 4 / (1 + \rho)^2 \right] P_{SW} \quad (5)$$

and for the travelling wave ( $TW$ ) case, when  $\rho = 0$ ,  $P_{TW} = 4P_{SW}$ .

**Multipacting in rectangular waveguides** appears between the surface of walls and is driven by the electric field. Therefore the analytic solution for parallel plates geometry holds.

Fig.3 shows multipactor susceptibility zones for parallel-plate geometry, aluminum waveguide test samples [3]. Horizontal scale: frequency [Ghz] · waveguide height [mm], vertical scale: peak gap voltage [V]. Dashed lines surround areas of possible multipacting zones (order  $n = 1$  at left, higher orders to the right). Multipacting zones are terminated at low values of the gap voltage because of a too low impact energy (secondary yield  $\delta < 1$ ). The solid line (lower right) separates multipactor free zones from those with possible multipacting. Circles indicate experimental observation of multipacting.

**Secondary electron yield** Multipacting occurs if there are resonant trajectories and if the secondary yield  $\delta > 1$ . Most metals will produce more than one secondary for one primary impinging electron. A typical curve of the secondary yield vs. impact energy is given in Fig.4 [4].

The secondary yield of Niobium and Cu plated steel can be substantially reduced by baking at 350°C. Tab.1, using the terms of Fig.4, shows the effect of baking in vacuum and of the loss of good effect upon reexposure to atmosphere

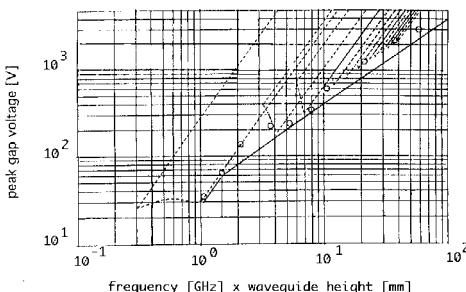


Figure 3: Multipacting in rectangular waveguides.

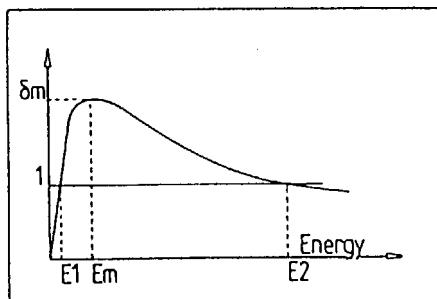


Figure 4: Typical curve of secondary yield of metal vs. impact energy (see Table 1).

Table 1: Secondary yield for Nb and Cu, and various processing.

| Material (process)                     | $\delta_m$ | $E_m$ [kV] | $E_1$ [kV] | $E_2$ [kV] |
|--|------------|------------|------------|------------|
| Nb (as received) [5]                   | 2.6        | 0.25       | 0.02       | >1.8       |
| Nb (300° C baked)                      | 1.4        | 0.25       | 0.05       | 1          |
| Cu (as received)* [5]                  | 2.5        | 0.25       | 0.02       | >2         |
| Cu (300° C baked)* [5]                 | 1.7        | 0.35       | 0.08       | >2         |
| Cu (cleaned)<br>(exposed 1 day to air) | 1.2        |            |            |            |
|  |            | 1.5        |            |            |

\* Cu plated onto stainless steel

[5]. The secondary yield of pure ceramics for rf windows is rather high (Sec.6.1) but can be reduced by appropriate coatings (Sec.6.15.1) such as Ti, TiN, CrO.

**Suppression of multipacting** There are several counter-measures helping to avoid multipacting: (i) Avoidance of resonant trajectories is the most effective means to eliminate multipacting. In the case of coaxial lines the resonance conditions scale with the fourth power of the distance between conductors. Therefore multipacting free operation can be assured by choosing the right dimensions of the coaxial line. One-point multipacting in sc cavities is eliminated by the use of an elliptical instead of the cylindrical shape. In this case the strong electric surface field destroys the magnetic driven trajectories [1]. Computer simulations of trajectories are very helpful in understanding and avoiding multipacting. There are several 2-D [2] and 3-D codes for multipacting simulation in accelerating structures and RF components, which solve non-relativistic equation of motion for electrons. Two recently developed codes, Track3P [6] and MultP-M [7], can model emission of electrons and their trajectories in 3-D space. Track3P uses parallel processing so it can

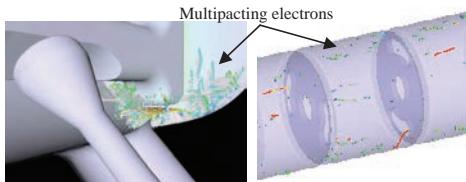


Figure 5: Modeling of multipacting with Track3P (left) and MultP-M (right).

scan over a wide parameter space in short time. It is based on finite element with conformal meshing for accurate modeling of the geometry. The code imports field maps from 3-D eigenvalue solvers like Omega3P. MultP-M imports 3-D fields from any text file. Figure 5 shows multipacting electrons in the filter of HOM coupler (left) and in a section of disk-loaded structure.

(ii) DC bias of a coaxial line suppresses multipacting [8, 9]. The right amplitude of a static electric field between the inner and outer coaxial line will repel electrons from the inner conductor (multipacting in a coax line is due to one point electrons from the outer conductor);

(iii) Reduction of the secondary yield by cleaning or coating techniques. The benefit of baking is described above. Rf processing is another way of cleaning the surface and reducing  $\delta$ . Reduction of  $\delta$  by coatings, such as mentioned earlier for ceramics windows, can be also effective. Klystron manufacturers sometimes sinter TaC onto drift tube tips to avoid multipactor in the gaps [10]. The draw-back of these methods is that contamination of the surface with gases during operation may initiate multipacting again;

(iv) Reduction of  $\delta$  by geometric means. Small grooves in the surface will “shadow” the lower surface against the electric field. Therefore only the “upper” surface can take part in the multipacting process. Simulations confirmed that the effective secondary yield is reduced by the ratio of the lower to upper surface. This method has been used successfully on coaxial lines for power tubes[10];

(v) Tapering of surfaces (coaxial lines e.g.) are reported to suppress multipacting in some cases. On the other hand first simulations of tapered coaxial lines predict a broad band of non resonant multipacting, because a taper will offer a band of distances for resonant trajectories;

(vi) An additional static magnetic field suppressed multipacting in normal conducting components [11].

## References

- [1] U. Klein, D. Proch, Proc. Conf. Future Possibilities for Electron Accelerators, N1-N17 (1979)
- [2] E. Somersalo et al, PA 59 (1997) 107
- [3] A. Wood, J. Petit, ESA Journal 1990, Vol.14, p.467
- [4] Handbook of Chemistry and Physics, 78th ed., CRC Press, p.12-125
- [5] R. Calder, G. Dominichini, N. Hilleret, NIM B13 (1986) 631
- [6] L. Ge et al., PAC09, Vancouver 2009
- [7] M. Gusarova et al., NIM PR A, 599, p. 100, 2009
- [8] P. Ylä-Oijala, DESY TESLA Report 97-21
- [9] J. Tucekmantel et al, PAC 95, p.1642
- [10] M.J. Smith, G. Phillips, Power Klystrons Today, Research Studies Press Ltd. (1995)
- [11] AERE TRANS 890, Harwell, Berksh., U.K. (1962)

## 6.10 POLYPHASE POWER CIRCUITS

*M. Tigner, Cornell U.*

(See Ref.[1])

Fig.1 shows a 3 phase source and load connected in a four wire, “Y” circuit, angle between phases being  $120^\circ$ .

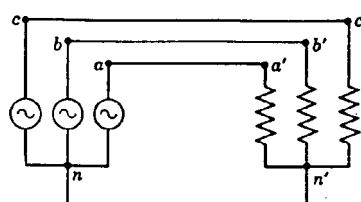


Figure 1: 3 phase, 4 wire circuit.

$a$ ,  $b$ , and  $c$  are referred to as the Line terminals and  $n$  as the neutral terminal. Thus the voltages  $V_{cb}$ ,  $V_{ba}$ ,  $V_{ac}$  are referred to as Line voltages,  $V_L$ , while the voltages,  $V_{an}$ ,  $V_{bn}$  and  $V_{cn}$  are referred to as Phase voltages  $V_P$ . Fig.2 shows the general vector diagram for the voltages and currents in such a circuit with unbalanced loads. Note that the neutral line current is not zero in this case of unbalanced loads. In 4 wire systems the neutral is often grounded for safety.

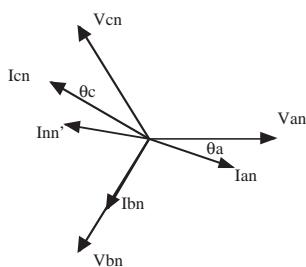


Figure 2: Voltage and current vectors.

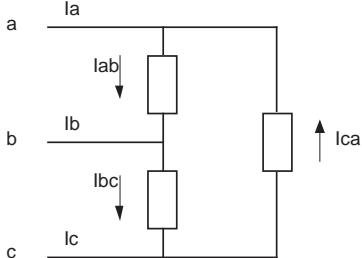


Figure 3: Delta connection.

In the usual convention,  $I_{an}$  is said to be lagging,  $I_{bn}$  to be of unity power factor and  $I_{cn}$  to be leading. The power in the circuit will be the sum of the powers dissipated in the three phases,

$$P_T = V_{an} I_{an} \cos \theta_a + V_{bn} I_{bn} \cos \theta_b + V_{cn} I_{cn} \cos \theta_c \quad (1)$$

In the usual case, the loads and voltages are balanced for proper operation. In that case

$$\begin{aligned} I_P &= I_L, \quad V_L = \sqrt{3} V_P \\ P_T &= 3 V_P I_P \cos \theta = \sqrt{3} V_L I_L \cos \theta \end{aligned} \quad (2)$$

and the power factor  $\cos \theta$  is the same for all phases. This power factor is the ratio between the magnitudes of the resistance and impedance of the load.

Alternatively, the so called  $\Delta$  connection may be used as shown in Fig.3. In this case the line and phase voltages are the same but

$$\begin{aligned} I_L &= \sqrt{3} I_P \\ P_T &= 3 V_P I_P \cos \theta = \sqrt{3} V_L I_L \cos \theta \end{aligned} \quad (3)$$

The three phase circuit is the most economical of polyphase power transmission circuits, requiring 25% less copper needed for the same power over the same distance, the same voltage between conductors and the same transmission loss.

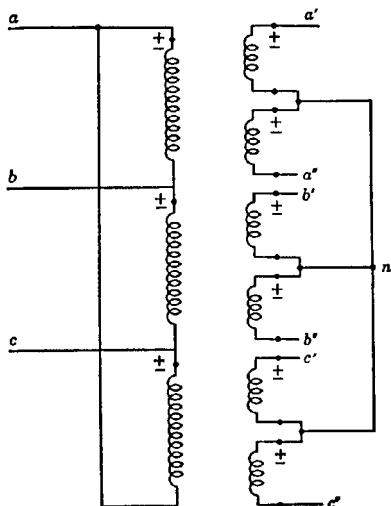


Figure 4: 6 phase secondary.

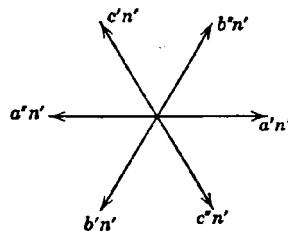


Figure 5: 6 phase vector diagram.

Nevertheless for some local end uses such as for rectifiers with low ripple, higher multiples  $N$ , are sometimes used, where the phase difference between lines is  $360^\circ/N$  with the exception of 2 phase circuits where the phases are in quadrature. Some examples of the line and phase relationships in common cases are given in the table below:

|     | $Y$   | $Y$           | $\Delta$      | $\Delta$ |
|-----|-------|---------------|---------------|----------|
| $N$ | $I_L$ | $V_L$         | $I_L$         | $V_L$    |
| 3   | $I_p$ | $\sqrt{3}V_p$ | $\sqrt{3}I_p$ | $V_p$    |
| 4   | $I_p$ | $\sqrt{2}V_p$ | $\sqrt{2}I_p$ | $V_p$    |
| 6   | $I_p$ | $V_p$         | $I_p$         | $V_p$    |
| 12  | $I_p$ | $0.52V_p$     | $0.52I_p$     | $V_p$    |

Figs.4 and 5 show a typical 6 phase transformer circuit.

## References

- [1] Jones, Laughton, Say, Electrical Engineers Reference Book, Butterworth-Heinemann (1996)

## 6.11 HIGH PRECISION POWER CONVERTERS

*F. Bordry, J.P. Burnet and  
M. Cerqueira Bastos, CERN*

### 6.11.1 Introduction to Magnet Power Converters

The trend in power electronics is evolving rapidly, mainly due to progress in power semiconductors. Since the nineties, the IGBT (Insulated Gate Bipolar Transistor) took the leadership in medium and large power, with the development of new converter topologies called switch-mode power converters [1], [2]. With the progress of electronics (analog and digital) and measurement systems, the performance of the power converters did a step forward in precision and stability. This chapter will give an overview of the state of the art for high precision power converters for particle accelerators (conventional and superconducting magnet, RF klystrons, solenoids).

### 6.11.2 Main Parameters of Magnet Power Converters

The powering of particle accelerator magnet are mainly DC power converters and in most of the case, the feedback control system regulates the magnet current. The choice of the power converter topology and then the technology will be defined by the required performance, translated in a precise specification. First of all the peak and the rms ratings of the current and voltage to power the magnet should be defined. Then, the operation in different quadrants has to be identified: shall the power converter be bipolar in current and/or in voltage? Shall the power converter operate as a generator and/or as a receptor? Fig.1 defines the 4-quadrants of operation of a power converter.

In quadrant 1 and 3, the power converter operates as a generator. In quadrant 2 and 4, the power converter operates as a receptor. In this case, the magnet is giving back its stored energy during the ramping down of the current. The power converter can dissipate this energy or it can give back this energy to the mains or it can store locally this energy which will be used for the next cycle. The power converter can be classified in 3 categories: one-quadrant power converter; unipolar in current and voltage, two- quadrant power converter; unipolar in current but bipolar in voltage, and 4- quadrant power converter; bipolar in current and

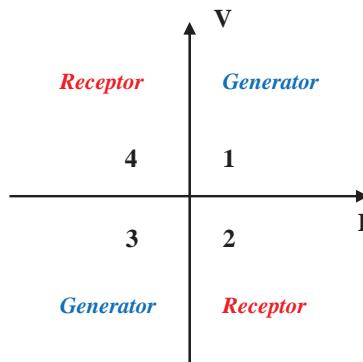


Figure 1: Definition of the 4-quadrant operation.

voltage. The power converter can be controlled with different strategies: steady DC current control, variable current reference, or pulsed current.

Another important parameter for the design of the power converters is the voltage and current ripple. The power converter topology and the performance of the inner control loops define the voltage ripple. The current ripple is defined by the load transfer function (cables, magnet inductance). To get good current ripple estimation, a good identification of the converter load is required.

To identify the optimal topology of the power converter, a complete list of parameters has to be reviewed between the accelerator physicists, the magnet designers and the power converter designers.

### 6.11.3 Power Converter Topologies

Three main families of power converter are used for particle accelerators: Thyristor-controlled rectifier, switch-mode power converter and discharged power converter. Each type will be described in the following paragraphs.

#### 6.11.3.1 Thyristor controlled rectifier

The thyristor-controlled rectifier was the main topology used from the seventies up to the nineties. Many different topologies can be done with thyristor devices, but only the three main types will be described here. First, the simplest one is the 6-pulse thyristor rectifier with free-wheeling diode, see Fig.1. The topology includes a transformer, a 6-pulse thyristor bridge with a free-wheeling diode and an output filter. The

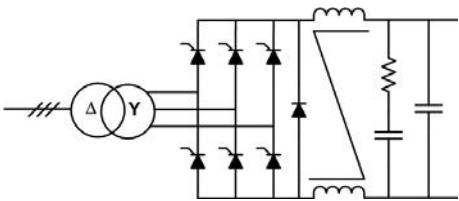


Figure 1: 6-pulse thyristor rectifier.

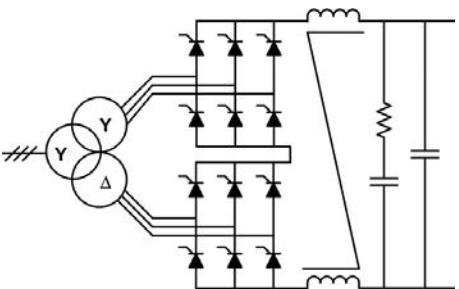


Figure 2: 12-pulses thyristor rectifier.

output voltage is controlled by the firing angle of the thyristor bridge. Due to the free-wheeling diode, the output voltage can only be positive, as well as the current. This topology is a one quadrant power converter. Second, the most used is the 12-pulse rectifier. This topology includes a transformer with two secondaries with a phase shift of  $\pm 15^\circ$ , two 6-pulse thyristor bridges connected in series, and an output filter, see Fig.2.

This topology allows getting an output voltage positive and negative but requires a minimum current to be able to control the thyristors. The first harmonic of the output voltage is 600Hz which helps to reduce the current ripple compared to the previous topology. When the voltage is negative, the energy is given back to the grid. It is a two-quadrant power converter.

Last, a back to back thyristor rectifier is shown as an example of 4-quadrant power converter. In this case, two 6-pulse rectifiers are connected back to back, see Fig.3. This topology requires placing inductors between the rectifiers to limit the circulating current which is necessary to control the thyristor rectifiers. In this case, the current can be positive and negative in the load as well as the voltage.

The thyristor devices are known to be robust, with low losses, and easy to drive. The main drawbacks of this type of power converters are

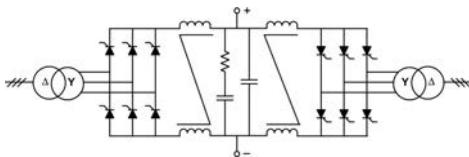


Figure 3: Back-to-back thyristor rectifier with circulating current.

to be very sensitive to the perturbations from the grid, it generates reactive power on the grid, and it has a low bandwidth. It also requires 50 Hz transformer which can be bulky. Nevertheless, this type of power converter is widely used for particle accelerators. It was for example, the main type for the LEP accelerator. Nowadays, the thyristor controlled rectifiers are mainly used for high power converter (above 500 kW) which requires very often a reactive power compensator on the grid.

### 6.11.3.2 Switch-mode power converter

Many types of switch-mode power converter exist and cannot all be listed in this chapter. Two main types will be presented, first, one topology with a 50Hz transformer and second, one topology with high frequency transformer.

**Switch-mode power converter with 50 Hz transformer** A transformer is always requested to insulate the magnet from the grid. If the volume of the power converter is not an issue, 50 Hz transformer can be used as it is widely produced by industry at low price. The most classical topology used for particle accelerators can be seen in Fig.4. The 50 Hz transformer adapts the voltage for the rectifier, then an H-bridge done with IGBT will control the voltage applied to the magnet. The switching frequency of the IGBT depends on the DC-link voltage and on the current to control. For power converters above 10 kW, the classical switching frequency starts from 1 kHz to 20 kHz.

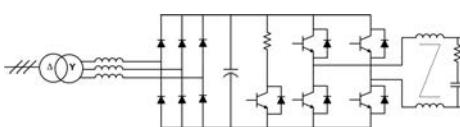


Figure 4: Switch-mode power converter with 50 Hz transformer.

The H-bridge topology allows to operate in the 4-quadrant. When the magnet gives back its energy, the return energy can be dissipate in a brake chopper or store in the capacitor bank of the DC-link. Another possibility is to replace the diode rectifier by an active front end done with IGBT. In this case, the energy can return to the grid. The main advantage of this topology is to use IGBT semiconductor which are widely produced at low price. They can be easily controlled with PWM (pulse-width modulation) technique at a switching frequency above few kHz which helps to reduce the current ripple and improve the feedback loop performance. The IGBT rating starts from 600 V up to 6.5 kV and from 50 A to 3 kA which allows covering a large scale of the needs.

**Switch-mode power converter with high frequency transformer** One of the main interests of using a high frequency transformer is to reduce the volume of the power converters. Ferrite cores are widely produced which allow using high frequency transformer at a competitive price. A classical topology can be found in Fig.5.

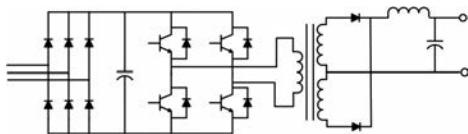


Figure 5: Switch-mode power converter with high frequency transformer.

The diode bridge is directly connected to the grids. An inverter drives a high frequency transformer. A diode bridge is connected at the output of the secondaries of the transformer to obtain a DC voltage to apply to the magnet. This type of power converter operates only in one-quadrant. 4-quadrant operation can be obtained by adding another stage but the return energy has to be dissipated at this level (Fig.6), [3].

The return of the energy to the grid can be obtained but with a much more sophisticated topology which will not be described here [1]. The

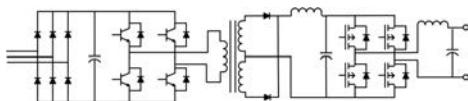


Figure 6: 4-quadrant switch-mode power converter with high frequency transformer.

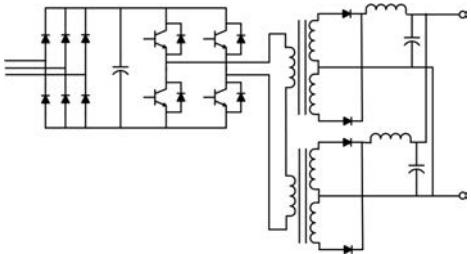


Figure 7: Switch-mode power converter with 2 high frequency transformers.

switching frequency of the inverter is usually from ten kHz to hundred of kHz. Due to the high frequency spectrum of the semiconductors commutation, an important design constraint for the power converters is the EMC (Electro Magnetic compatibility) immunity and emission. Soft commutation of the inverter semiconductors is an elegant solution to reduce the switching losses and to improve the EMC. In the case of LHC where power converters have to be installed underground with a limited space, the switch-mode power converter was chosen also for power converter up to 200 kW. In this case, many sub-converters have to be placed in parallel to reach this power level. The superconducting magnets require a high current (many kA) but few volts (less than 20 V); the solution was to drive in series many high-frequency transformers with one inverter. All the secondaries were connected in parallel after the output filter, see Fig.7.

To implement a redundancy policy, N+1 sub-converters were placed in parallel, where a sub-converter is made with one inverter with 8 transformers. In case of failure of one sub-converter, the other sub-converters will compensate without any disturbance to the load. A modulator construction of the power converter can be done due to the high number of converters in parallel which eases the operation and maintenance of it. For power converter below 1 kW, the best semiconductor is the MOSFET. The advantage of the MOSFET is to have lower losses than the IGBT which allows a higher switching frequency. Nevertheless, its voltage range is limited from 5 V to 500 V.

**Fast pulsed power converter** When the presence of the magnetic field in the magnet is needed for a very short time, discharged power converter is a very interesting technique to reduce power

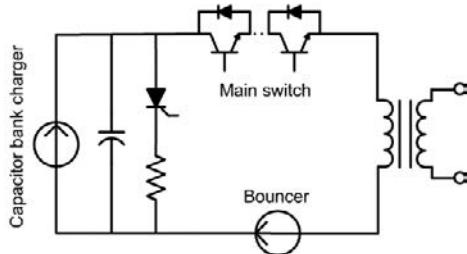


Figure 8: Klystron modulator with pulse transformer.

consumption. This is the case for example for beam transfer line where the beam can be present few s every second or minute. It is also the case for Klystron Modulator in a Linac. An example of discharged power converter topology is shown in Fig.8.

In this case, a capacitor bank is charged at a nominal value (less than 15 kV) which can be low compare to the need of the load (from 60 kV to 140 kV). A main switch discharges the capacitor bank via a pulse transformer.

The voltage is applied to the load as long as the switch stays ON but generally for a short time (2.4 ms for LINAC4), as the capacitor bank voltage decreases (few %), the droop voltage has to be compensated by a bouncer which is a passive resonant circuit. The main challenge is to design the pulse transformer and the main switch. In case of arc in the Klystron, the energy deposit has to be limited to few joules ( $\sim 20$  J). This needs a fast turn OFF of the main switch. A redundancy policy has to be implemented to be sure to be able to open the circuit. For LINAC4, 4 IGCT (Integrated Gate-Commuted thyristor) are placed in series where only 3 are required.

**High power system with local energy storage**  
For high power system (above 1 MW), Thyristor rectifiers are the main solution. The principal drawbacks are:

- the reactive power generated on the grid which needs to be compensated to stabilize the network voltage
- the pulsed active power on the grid which requires a strong electrical network. To avoid the flow of pulsed power on the grid, local energy storage can be used.

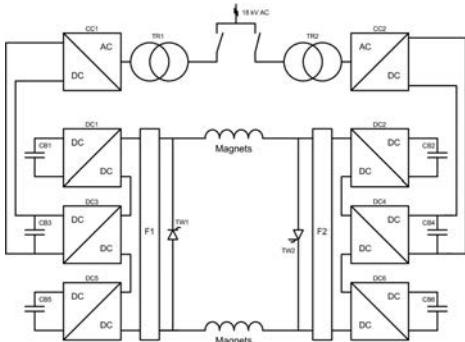


Figure 9: Power system with capacitive energy storage.

One example is the new POPS system [4] designed for the CERN PS accelerator, see Fig.9.

The principle is to store energy in capacitor banks and to exchange this energy with the magnets during the cycles. As the rating power is very high (60 MW peak), many switch-mode power converters are associated in series and parallel. Only the losses of the system (magnets and converters) are taken from the electrical network (5 MW peak). 20 MJ are stored in capacitor banks where 14 MJ can be exchanged during each cycle with the magnets. A redundancy is also implemented; the system can operate without one of the main components. In this case, no reactive power is generated on the grid and the active power taken on the grid is reduced to the minimum. This new type of power system will help the particle accelerators to reduce their energy consumption.

## References

- [1] F. Bordry, 11th European Conference on Power Electronics EPE2005, Dresden (Germany), September 2005.
- [2] F. Bordry, Power converters: definitions, classification and converter topologies, Specialised CERN Accelerator course on "Power Converters" -Warrington, UK -May 2004
- [3] A. Beuret et al, 12th European Conference on Power Electronics and Applications, Aalborg, Denmark, September 2007
- [4] F. Bordry et al, EPE Journal, Vol. 18, no 4, December 2008

## 6.12 HIGH ACCURACY IN POWER CONVERTERS FOR PARTICLE ACCELERATORS

*F. Bordry, J.P. Burnet and  
M. Cerqueira Bastos, CERN*

Magnets in particle accelerators are powered by electrical power converters. These can be seen as controlled current sources in which a feedback loop ensures that the output current follows the reference to the best possible accuracy. The accuracy that can be achieved is greatly determined by the current measurement transducer and, in the case of digitally controlled power converters, the control algorithm and the ADC employed in the feedback loop.

As energies reached by particle accelerators become bigger, accuracy requirements for the electric current which produces the magnetic field that steers the particle beam, increase proportionally. Accuracy requirements depend also on the powering strategy chosen for the accelerator. In a synchrotron, main dipole and quadrupole circuits are normally powered in series to ensure synchronism and homogeneity in the magnetic field around the circumference. However this is not always true: in the case of the LHC the main circuits are divided into eight sectors due to the very high stored magnetic energy and constraints in the protection of the superconducting magnets. As a consequence, individual sector currents need to be controlled with very high absolute accuracy in amplitude and time [1] to ensure tracking between all sectors. Requirements for accuracy in current control are also determined by the type of magnet and its function: the need for accuracy in the current control for corrector magnets is much less stringent than for quadrupole and dipole magnets.

### 6.12.1 Power Converter Control

In the past, power converter control in accelerators was usually based on analogue feedback loops using PID (Proportional/ Integral/ Derivative) control. In such systems, the reference value at the input of the loop is often given by a DAC (Digital to Analogue Converter) and the output current of the power converter measured by a precision current transducer which provides the feedback signal for the control loop [2]. The accuracy achieved with such techniques is limited by errors due to drift, linearity and temperature dependency. Moreover, adjustment of control parame-

ters can be cumbersome, as it might require trimming of potentiometer or replacement of components.

Developments in digital electronics and in particular DSPs, PLDs and Microcontrollers, as well as the need for increased performance in power converter applications fuelled significant progresses in digital control during the last decades. The first applications using digital control in power converters for accelerators were implemented in late 1980s [3]. Amongst the advantages of using digital control are increased stability and reproducibility, less susceptibility to noise and thermal effects, easy implementation of different control methods (state-space, robust, fuzzy) as well as easy loop parameterization. On the other hand, the use of digital control increases system complexity and introduces new sources of error such as the ones resulting from ADC measurement uncertainty and limited resolution on arithmetic calculations, which might lead to arithmetic errors.

When high accuracy is required, one effective power converter control strategy is to have an external current loop controlling a power supply working as a voltage source. In a digitally controlled system, the output current of the converter is read by an ADC connected to a current transducer and then compared with a digital reference. The error is fed into a digital regulator and the result sent to a DAC that provides an analogue signal to control the voltage source.

This solution is implemented in the control of the LHC power converters at CERN. In the case of the LHC, the control challenge is even bigger due to the 8-sector powering strategy used for the main dipole and quadrupole circuits. In this case, not only the control of each converter has to be extremely accurate but also the generation of the different current references and their synchronization along the 27 km circumference of the LHC. For this purpose each power converter in the LHC has a dedicated controls electronics which is actually an embedded microcontroller-based computer capable of performing full local state control, reference function generation and measurement acquisition as well as running a digital current regulation loop. Reference functions are synchronized using a timing network. Each digital controller is connected to a field bus (WorldFIP) and the timing network is used to synchronize the cycles of all segments of the field bus. The

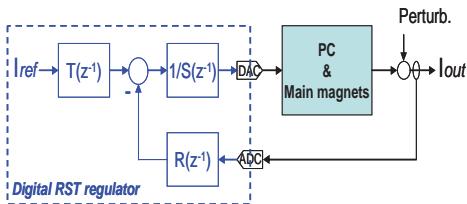


Figure 1: Digital RST control, used in the LHC at CERN.

digital controller disciplines a phase-locked loop to align its clock to the start of each WorldFIP cycle guaranteeing in this way synchronism of the references along the machine.

The digital control strategy implemented in the LHC power converters is based on an RST algorithm. This type of current loop (Fig.1) has the advantage of decoupling tracking behaviour (following of a reference) from regulation (disturbance rejection) [4]. Recent results proved that the tracking error on the LHC main power converters is in the order of one ppm and that tracking between different sectors is within a couple of ppm [5].

### 6.12.2 Current Measurement in Particle Accelerators

Traditionally, accurate current measurement devices for particle accelerators are associated with beam current measurement. Magnetic transducers and in particular current transformers have been for a long time the preferred transducers to measure beam currents. The requirements for beam current measurement have driven progress in transducer technology, culminating with the invention of the DCCT (Direct-Current Current Transformer) at CERN in late 1960s. The idea was to build a magnetic beam current transformer with frequency response extended down to DC to measure beam current in the ISR accelerator. The new transducer combined the zero flux detection principle used in flux-gate magnetometers since the 1930s, with the active transformer circuit as originally proposed by H.G. Hereward (and already used to measure the circulating beam in the CERN PS accelerator). Although the new transducer was not initially intended for power supply regulation applications, its advantages compared with previous DC instrument transformers (e.g. Kramer and Hingorani [6]) soon became obvious. The concept got picked up by industry and

the new transducer was soon being used at accelerators such as DESY in Hamburg. In late 1970s, DCCTs were used for the first time in large quantities in the SPS project at CERN [7].

The use of DCCTs spread to other applications but it continued to be widely used in particle accelerators. In particular, this last decade saw important progress in DCCT technology with the development and deployment of the DCCTs for the main dipole and quadrupole power supplies of the LHC, at CERN. Short term stability in the order of two part-per-million (ppm), yearly drifts better than fifteen ppm and linearity better than two ppm have been achieved.

**Current measurement technologies** The most common current measurement technologies used in electrical power converters include resistive shunts and current sense resistors, Hall-Effect current transducers (based on the polarization of charges in an electrical conductor in the presence of an external magnetic field), Current Transformers, Rogowsky Coils (high current, high bandwidth applications), Active CTs and DCCTs (both based on the zero flux detection principle).

The choice of a current measuring device for a specific application depends on factors such as current range, bandwidth, required accuracy, required output signal, need for isolation, reliability, installation constraints, availability and cost. DCCTs provide isolated measurements for different current ranges and can reach very high accuracy, albeit with a higher cost.

**DCCTs** A DCCT is a magnetic current transducer of the zero-flux type where a second harmonic or a peak detector is used in a feedback loop to generate a compensation current which is a fractional image of the current being measured. (Fig.1)

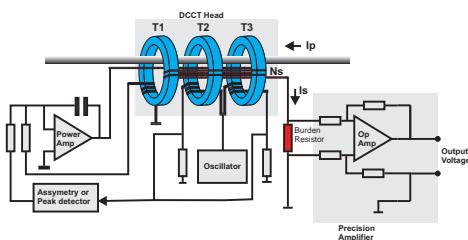


Figure 1: The DCCT working principle.

DCCTs have the potential of reaching short term stability and repeatability in the order of a few part-per-million. This requires not only a careful design of the magnetic part (magnetic head) as well as the use of high quality burden resistors and very stable precision amplifiers, usually in a temperature controlled environment.

The head design and in particular the magnetic shielding are important to increase the sensitivity to primary magnetic field while minimising sensitivity to external magnetic fields, head centring and return bus-bar fields. Soft magnetic materials and in particular amorphous and nanocrystalline alloys are the most commonly used in DCCT core and inner shielding design because of their magnetic properties such as high initial permeability and low coercivity. As for the outer layer of the shielding, usually a doughnut-shaped shell containing the cores and the inner shielding, ferromagnetic materials of lower permeability and high saturation are normally used. During the last couple of decades, improvements in the manufacturing process of amorphous alloys contributed to their widespread use and to the development of new materials hence making for a greater availability of candidate materials for magnetic sensors. More recent developments include progresses in reducing coercivity and increasing saturation induction of amorphous and nanocrystalline materials. These new breakthroughs have still to find their application in DCCT head design (Fig.2).

The current output of the DCCT is usually connected to a burden resistor, a 2-terminal or 4-terminal resistor depending on the required accuracy. The performance of this component is one

of the dominant factors in the overall precision of the transducer. Well known effects that can cause resistance change and therefore can affect the performance of DCCTs are temperature coefficient, self heating and thermal settling as well as ageing. Less well known effects include power coefficient, humidity absorption and hysteresis under power cycling. The highest accuracy available in burden resistors known at the moment is offered by a proprietary Zeranin wire design from HITEC Power Protection, which is used in the LHC main dipole and quadrupole DCCTs, but its price limits its use only to the best DCCTs. Otherwise there is only one resistor type on the market offering the performance needed: Bulk Metal Foil or foil. This technique, pioneered by Vishay, but now widely spread, tightly bonds a rolled metal foil to a substrate and seeks to compensate the resulting consistent stress effects as part of the overall resistor performance [2].

The voltage across the burden resistor normally needs to be amplified to produce a voltage output adequate for subsequent ADC conversion. This task is done by a precision amplifier, usually a difference amplifier circuit making use of high precision network ratio resistors to establish the gain. Presently, the consumer electronics market offers a range of choice in precision amplifiers, including low offset, low drift amplifiers that can be used as the output amplifiers in the DCCT. However, this does not exempt the designer of careful design and implementation of the output circuit. Parameters such as offset and gain stability, noise, common mode rejection and output impedance depend strongly on circuit design and implementation.

**ADCs** With the advent of digital control, Analogue to Digital Converters have become essential components for ensuring the accuracy of the current measuring chain in the control loop of power converters. If in the past, most of the control was implemented using analogue electronics, in the last couple of decades digital control has almost completely taken over. The advantages of having a digital representation of the signal are numerous and not limited to power converter control: calibrations, traditionally done by adjusting gain and offset potentiometers, is now performed using calibration constants which can be memorised and used to correct the output of the measuring device. If these constants are also kept in a database and updated whenever a calibration takes place,



Figure 2: A 13 kA DCCT head, used in the LHC.

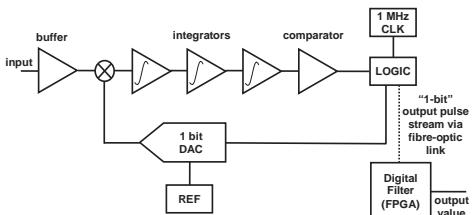


Figure 3: The Delta-Sigma Analogue to Digital converter used in the LHC main power converters.

the task of following the behaviour of high precision devices becomes much simpler and less error prone. Another advantage is the possibility of using correction algorithms. A typical example is the use of a temperature sensor and a correction algorithm to correct for the temperature dependency of a measurement device.

ADC technology has significantly evolved in the last couple of decades fuelled mainly by the telecommunications industry. As a consequence, a wide range of solutions in AD conversion are available in the market with tradeoffs in resolution, speed and accuracy. In power converter control for accelerators, ADC speeds above the MHz are seldom required. As for resolution, requirements often range from 16 bits to 24 bits with accuracy following along. In this context Successive Approximation Register ADCs are one of the most interesting technologies as they cover the medium speed, medium accuracy range, with resolutions up to 18 bits. Integrating ADCs can also be used in the medium accuracy (albeit low speed) range for resolutions up to 18 bits. Although integrating ADCs are well suited for precision applications, they have the disadvantage of requiring precision external components. When high accuracy and high resolution (18 to 24 bits) are required, Delta Sigma ADCs are the preferred solution (Fig.3). They dont require external components and their oversampled nature simplifies circuit design by relaxing the requirements of the analogue anti-aliasing filter. However, higher accuracy also requires more complex digital decimation filters which, for linear phase FIR filters, corresponds to increased latency. Filter latency limits the maximum loop bandwidth which means that high accuracy is normally associated with lower speed. The use of minimum phase FIR filters can substantially decrease latency.

ADC accuracy depends greatly on the voltage reference employed. Precision external voltage references are preferred in applications demanding high accuracy as they have lower temperature coefficient, thermal hysteresis and long term drift than an on-chip voltage reference. Zener based high precision references are usually referred to as Buried Zener references. They can include internal temperature control or temperature compensation. However, for higher accuracy, they might need to be stabilised inside a temperature controlled oven. Reference annealing can also be used to accelerate the ageing process and reduce initial drift.

The main dipole and quadrupole power converters on the LHC use a Delta Sigma converter with a resolution of about 22 bits. The voltage reference is a buried zener type, previously annealed for minimum drift. Sub-ppm accuracy is achieved albeit at very low sampling speeds (1 kHz) and in a temperature controlled environment.

**Calibration** There can be different motivations for calibration of measurement devices in accelerator applications. The first is the requirement to keep long term drift within certain limits. Another motivation is to minimise the impact of replacements, i.e. when a device in operation is replaced by another one that has been sitting on a shelf, if they havent both been calibrated, there might be a significant change in the operating point of the converter. Finally, calibration might be necessary to guarantee good tracking between different transducers. Such is the case in the LHC at CERN, where the main dipole and quadrupole circuits are divided into eight sectors. In this case, calibration of the current measurement chain using the same reference ensures that all sectors track each other, allowing the beam to circulate around the machine without seeing any difference in the magnetic field between sectors.

The periodicity of calibrations must be set according to the accuracy requirements of the power converter. Some devices include automatic calibration mechanisms: it is common to calibrate ADCs using a multiplexer that connects the ADC input to a voltage reference for calibration. However, the references themselves might need periodic calibration, so human intervention can be unavoidable.

Calibration procedures usually involve the use of dedicated equipment which must be previously characterised in a laboratory by using well

known reference devices. The complexity of the necessary calibration infrastructure depends on the accuracy trying to be achieved. Some of the methods employed to calibrate DCCTs and ADCs and required calibration equipment are described below.

DCCTs can be calibrated through different methods:

- The reference DCCT method, where the output of the DCCT being calibrated is compared against the output of a reference DCCT measuring the same current.
- The output stage method, which consists in injecting a reference current in the burden resistor of the DCCT and measuring its output with a calibrated ADC or with a DVM. The value of the error is then memorised by the digital controller. This method has the disadvantage that it only calibrates the DCCT output stage (burden resistor + precision amplifier) and it requires a precision current source to generate the calibration current.
- The calibration winding method, which consists in injecting a reference current in an auxiliary winding in order to produce an Ampere-Turn value equivalent to the one produced by the primary current hence simulating real primary current. The output of the DCCT is measured with a calibrated ADC or with a DVM and the value of the error is memorised by the digital controller.

The latter method requires a precision current source to generate the calibration current. At CERN, a programmable current reference has been developed for the calibration of DCCTs equipped with calibration windings. It can produce DC currents ranging from -5 A to 5 A with sub-ppm uncertainty. In the LHC, in-situ calibration systems equipped with these devices are installed close to the main power converters (Fig.4). They are housed in temperature controlled racks and can be remotely controlled via an ethernet connection in order to calibrate one or several DCCTs in the main dipole and quadrupole power converters.

In other applications, such a current source might not available. In such cases, DCCTs are usually calibrated using the reference DCCT method. However since in-situ deployment of a reference DCCT is not always easy, the procedure

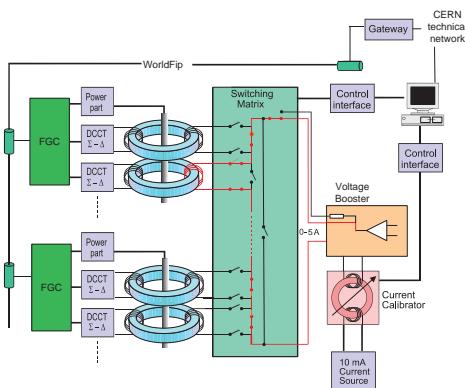


Figure 4: The remotely controlled calibration system for the main power converters in the LHC.

often requires bringing the DCCT to a laboratory. Usually only the DCCT electronics are removed from the converter since removing the heads is a difficult process. Calibration in the laboratory is therefore performed using a different magnetic head connected to the electronics, which can introduce errors.

ADCs can be calibrated using different methods depending on the parameters being calibrated (offset, gain, linearity). The most common is a three point calibration i.e. calibrate at FS and at zero. Calibration at FS can be performed using a voltage standard connected to the input of the ADC. The use of voltage standards requires constant monitoring and maintenance of these devices, including periodic calibration in an independent standards laboratory. In the LHC case, a set of voltage, resistor and even some unique current standards, are kept in a controlled environment to be used as the basis for CERN's calibration infrastructure. Some units are foreseen for field calibration whilst a set of standards is kept constantly in the laboratory, under permanent monitoring.

## References

- [1] F. Bordry, Regulation Theory, Specialised CERN Accelerator course on "Power Converters" - Warrington, UK -May 2004
- [2] I. Barnet et al, CERN, Geneva, Switzerland, 5th European Particle Accelerator Conference, Sitges, Barcelona, Spain, 10 - 14 Jun 1996, 2317
- [3] J. Carwardine, F. Lenkszus, ICAL-EPCS99, Trieste, Italy, (10/4/99 - 10/8/99) (2000) 171

- [4] F. Bordry, H. Thiesen, Electrical power Techonology in European Physics Research EP2, Grenoble (France), Oct. 1998
- [5] J. Carwardine, F. Lenkszus, Journal of Magnetism and Magnetic Materials, Volume 304, Issue 2, September 2006, 187
- [6] C. Adamson, N.G. Hingorani, Proceedings IEE, Vol. 110, No 4, April 1963, 739
- [7] H. Appelo, M. Groenboom, J. Lisser, IEEE Trans. Nucl. Science, Vol NS-24, No. 3, June 1977, 1810

## 6.13 HIGH-GRADIENT LIMITATIONS IN ROOM TEMPERATURE RF LINACS

*G.A. Loew, S. Tantawi and J.W. Wang,  
SLAC*

The design of high-gradient room-temperature rf linear accelerators requires that special attention be paid to several phenomena which limit their performance: electron field emission, dark current, pulse heating, rf breakdown and cavity damage. These are all discussed in this section. The reader should note that this is a field in progress, and that new experiments which have been performed since the previous printing of this Handbook have led the authors to make some significant changes. (For superconducting linear accelerators see Secs. 7.3.9,7.3.10,7.3.11.).

**Field emission** Despite the new results mentioned above, electron field emission (FE) [1] is still believed to play a central role in limiting high-gradient operation of rf linacs. FE was first calculated for dc fields. It results from the quantum mechanical tunneling of conduction electrons through a modified potential barrier at an ideal, perfectly clean, flat metal surface in an applied electric field. The FE dc current density in A/m<sup>2</sup> is generally expressed as (see also Eq.9, Sec.6.14)

$$j_{FE} = \frac{1.54 \times 10^{-6} \times 10^{4.52\phi^{-0.5}} E^2 e^{(-k\phi^{1.5}/E)}}{\phi} \quad (1)$$

where  $k = 6.53 \times 10^9$ ,  $E$  is the externally applied surface field in V/m and  $\phi$  is the work function of the metal in eV. For practical metallic surfaces (typically copper), the field that must be used in Eq.(1) is the microscopic local field which is obtained by multiplying the theoretical surface field by an enhancement factor  $\beta$ . Enhancement (i.e. increase) of the surface field can be due to

a number of causes: metallic protrusions, metallic or dielectric impurities inherent in the material or resulting from fabrication and cleaning processes, dust, and grain boundary imperfections, all of which are now commonly called "emitters". Although there is no reliable method to predict the value for a given emitter, it can be derived from empirical methods. Eq.(1) has to be modified somewhat in the case of rf fields. If we assume that the emitted current is generated from an area  $A_e$  (resulting from a single dominant emitter or a collection of emitters of similar strengths), then the emitted rms current in A is [2]

$$\bar{I}_{FE} = \frac{5.7 \times 10^{-12} \times 10^{4.52\phi^{-0.5}} A_e (\beta E)^{2.5}}{\phi^{1.75}} e^{\left(\frac{-k\phi^{1.5}}{\beta E}\right)} \quad (2)$$

$\beta$  can be determined from the slope of a semi-log plot of  $\bar{I}_{FE}/E^{2.5}$  vs  $1/E$  through

$$\frac{d [\log_{10} (\bar{I}_{FE}/E^{2.5})]}{d(1/E)} = -\frac{2.84 \times 10^9 \phi^{1.5}}{\beta} \quad (3)$$

A large number of experiments have been done over the past years which give typical values for  $\beta$  in the range 40-100. These values as well as the absolute values of  $\bar{I}_{FE}$  can change as a function of time during rf processing of a given rf structure, indicating that the quality of the surface and the properties of the emitter(s) are modified. Interestingly, after many hours of processing and breakdown, the metallic microprotrusions observed with a scanning electron microscope exhibit ragged shapes (craters and drops of molten copper) whose geometry can at best explain  $\beta$ 's between 5 and 10, unless one assumes multiplicative effects of so-called "tip-on-tip" protrusions ( $\sim \beta_1 \beta_2$ ) or other surface physics phenomena such as the presence of absorbed gas, dielectric layers, metal-insulator-vacuum (MIV) or metal-insulator-metal (MIM) effects [3]-[9]. In practice, the FE current generally becomes measurable (on the order of a few  $\mu$ A peak) when the local microscopic field exceeds a few GV/m. With a typical  $\beta$  of 50, this corresponds to an applied macroscopic surface field of roughly 40 MV/m to a perfectly flat surface. In a common room-temperature disk-loaded accelerator structure, the TW accelerating field on-axis is down by a factor of about 2 from the peak field on the disks, or 20 MV/m. However, because of the exponential growth of  $\bar{I}_{FE}$  vs.  $E$ , this current increases very rapidly with gradients above these values.



Figure 1: Prototype accelerator structures for the NLC/GLC main linac. RDDS stands for Rounded Damped Detuned Structure. HOM stands for Higher Order Mode.

**Dark current** When FE electrons are first generated in an accelerator structure, they come preferentially from high-field regions such as disk edges. At relatively low gradients, their trajectories and the points where they hit a cavity wall are fairly random because of their low emitted energy. Together with secondary electrons emerging from impacted areas, they end up producing the equivalent of an electron gas in each cavity (See Fig.13 in [10]). As the accelerating gradient in a velocity-of-light structure is raised [11] above a critical value  $E_c$  given by  $1.6/\lambda$  MV/m where  $\lambda$  is the rf wavelength in meters, axial capture and cumulative acceleration occurs. One problem with this captured current is that it can build up along each structure and from structure to structure. Captured electrons parasitically absorb rf energy, cause increased radiation, backgrounds, noise in instruments, and possibly wake-fields. Fortunately, there is evidence [12] that the captured electrons are to a large extent deflected transversely and eliminated by the usual focusing quadrupole array along the linac (whose strength scales with length, unlike the energy of the locally captured particles). Hence, the risk of these electrons reaching the end of the linac and the collision point is considerably decreased, whereas local radiation, noise and possibly damage remain a problem. Clearly, the  $1/\lambda$  dependence of  $E_c$  favors high rf frequency linacs in order to minimize capture.

**Pulse heating and direct cavity damage** Whereas normal FE can be a stable, steady-state

phenomenon, rf breakdown is an instability which can lead to a “runaway” condition and structure damage. While it was believed in the past that FE is the initial cause of rf breakdown, recent experiments cast at least some doubt on this belief. Although the entire mechanism is not yet understood, it appears that pulse heating, somewhere in the cavity where the magnetic field is high, may be the original cause. Although the entire mechanism is not yet understood, evidence presented in the next section indicates that pulse heating, somewhere in the cavity where the magnetic field is high, may be the original cause.

Note that in most internal walls of pulsed rf accelerator structures made out of copper (or other metal), the magnetic fields and the corresponding electric currents are fairly low and heating of the surface does not cause physical problems. However, in the neighborhood of irises or disk slot edges (used to damp higher-order modes), these magnetic fields and currents can theoretically cause initial heating exceeding 100° C within a pulse [13], [14], [15], [16]. A practical formula to estimate the temperature increase  $\Delta T$  is given by:

$$\Delta T = \frac{H_t^2}{\sigma \delta} \sqrt{\frac{t}{\pi \rho C_e \kappa}} \quad (4)$$

where  $H_t$  is the surface magnetic field,  $t$  is the time into the pulse,  $\sigma$  is the electric conductivity,  $\delta$  is the skin depth,  $\rho$  is the density,  $C_e$  is the specific heat, and  $\kappa$  is the thermal conductivity of the metal.

In extreme cases, this heating can induce thermal fatigue and thermal stresses beyond the elastic limit of the metal. After some millions of pulses, these are observed to produce surface roughness and cracking, and further enhanced pulsed heating. The threshold for the damage due to pulse heating is still unknown but the calculated temperature increase of over 100° C for an initially undamaged surface gives an experimentally observed warning point not to be exceeded.

**Pulse heating as a possible trigger for rf breakdown** The above observations made in tests of X-band (11.424GHz) accelerator structures have led designers to round off sharp edges in matching irises, coupling irises and disk slots to practical levels [see Fig.1]. For predicted temperature rises of less than 20° and 50° C respectively, these improved structures have performed without any pulse heating damage. Note that these modifications have in no way affected the accelerating gradients of the structures.

But even before these limits are reached, it appears that the onset of pulse heating somewhere in the cavities can lead to electrical breakdown somewhere else in the cavities, namely at the high electric field points at the irises. The reason for this conclusion is based on the following observations. Three short disk-loaded standing-wave structures with respective  $a/\lambda$  values of 1) 0.105, 2) 0.143 and 3) 0.215 were built and tested [17]. Surprisingly, as seen in the plots shown in Figs.2 below, the probabilities of breakdown did not correlate with the gradient (a) or the peak electric field (b), but with the peak magnetic field in the cavity (c). This new empirical result does not yet provide a complete mechanistic explanation of the breakdown but seems to imply that pulse heating at a high magnetic field point in the cavity causes some disturbance which then triggers an electric breakdown on the disk edges during the same or a subsequent rf pulse. More research work is needed.

**Final rf breakdown and cavity damage** Whatever the initial trigger for breakdown is, the exact sequence of events between initial rf processing (predominantly outgassing) and ultimate processing when structures are pushed to their limit is not totally clear. At a microscopic field of several  $\text{GV/m}$ ,  $j_{FE}$  can approach  $10^{11} \text{ A/m}^2$ . At this level and above, the  $ri^2$  Joule losses in an emitter can heat it in a few or tens of nanoseconds to the point of vaporizing some of its contents (gas, metal,

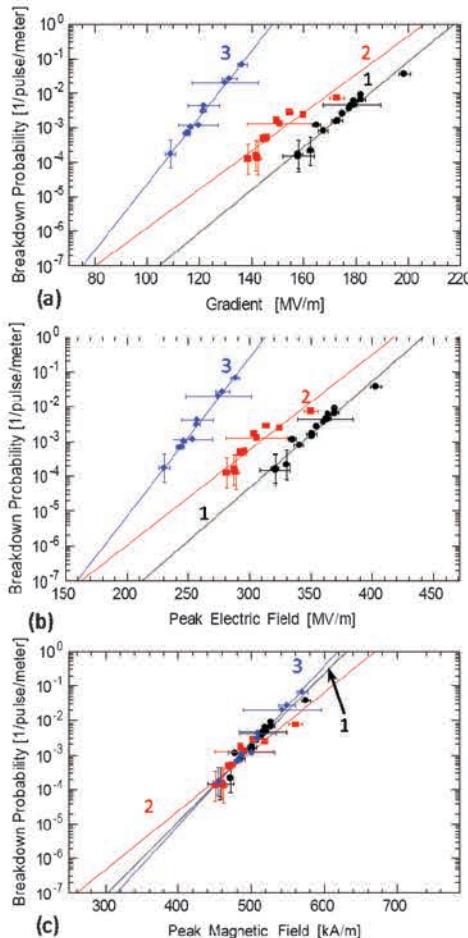


Figure 2: Comparison of breakdown probabilities for three different structures. Their respective  $a/\lambda$  values are given earlier in the text.

etc.) [4], [10], [18]. This process can be beneficial by smoothing and cleaning the surface. However, the positive ions, produced by ionization of the neutral gas by the FE electrons, can form a sheath or plasma spot [19], [20] over the emitter, precipitating further increased FE, more ions, eventual explosion or meltdown, metal droplets, and expanding damage in the form of craters. Experimental evidence for these phenomena is corroborated by the well documented observation of visible sparks, instantaneous increase of the emitted current by a factor of  $\sim 40$ , detection of puffs of gaseous H, CH<sub>4</sub>, H<sub>2</sub>O, CO and CO<sub>2</sub>, x-ray bursts,

and collapse (within the RF pulse) of the structure fields, manifested by a reflection of the incident RF power and/or a sharp decrease in the transmitted field [10], [21].

Experimental data obtained from early tests with a variety of single copper cavities and two short resonant accelerator sections [10], [20] indicated that the maximum surface fields,  $E_s$ , attainable after aggressive rf processing, i.e., where the incident power was driven up without concern of producing surface melting and craters, were  $\sim 310$  MV/m at 2.856 GHz and 500 MV/m at 11.424 GHz. However, the resulting surface damage would have been unacceptable for a practical accelerator structure because the craters severely detuned the operating frequency of the cavities. In the last ten years or so, the drive to build reliable high-gradient structures for electron-positron colliders led to a very large number of room-temperature structure designs and tests, mostly at 11.424 GHz and some at around 30 GHz which will be described below.

**Practical structure designs, RF processing and other criteria** We will now summarize observations and results which have been obtained from experiments performed in recent years for the design of high-gradient structures for linear colliders. The tests of structures for linear colliders have been done predominantly at 11.424 MHz, with only a few others at 30 GHz. All the X-band structures, as will be seen below, reached maximum practical gradients in the range of 60 to 80 MV/m after considerable rf processing.

Breakdown in these structures, once they have reached a desired level, results in two problems. The first is that during the pulse(s) when breakdown occurs, the structure is unlikely to produce acceleration since power transmission is interrupted at some point. When this happens, the total energy of the accelerator will fluctuate in an unacceptable way. The second problem is that breakdown produces structure damages in the form of craters and crystal boundary cracks. If too many such damages are created, the frequency of the cavities where they occur will change. This detuning changes the phase velocity and reduces the net acceleration because the synchronism between the electron or positron bunches and the wave is affected.

The first problem is caused by imperfections in the copper (or other metal) surfaces, impurities, inadequate baking, incomplete thermal or

rf processing, and by operating above thresholds which are not fully understood or predictable at this time. For example, all the X-band structures were treated with the following consecutive steps: accelerator cup surface light etching and cleaning, clean room assembly, wet and dry hydrogen firing and extended (two weeks) high temperature ( $650^\circ$  C) vacuum baking.

The second problem is not completely understood either. This observation reduced the available design parameter space because in constant-gradient structures where the group velocity is higher at the input and lower at the output, the higher power near the input is more likely to cause damage. As a result, structures have to be shorter (thereby increasing the total number of feeds per unit length) and must have lower group velocities (both at the input and the output) to reduce the probability of breakdown and damage. In practice, for a given input power level, the group velocity ranges where damage could be avoided were between 4% and 1% of the velocity of light. In order to meet the above requirements of length, group velocity range and optimized structure rf efficiency, a design with higher phase advance per cell ( $5\pi/6$  mode for SLAC case) had to be used.

The figures shown below [22] summarize all these experiments at X-band at SLAC. The recipe for successful processing is to carefully process each structure to an asymptotic field level above which no more improvement is achieved within a practical time, but below the level where damage becomes inevitable. Another criterion which must be met is that for a given collider design, the number of allowable pulses lost per unit time due to breakdown is limited. Typically, for the NLC structures discussed here, this allowable number is below the green dotted horizontal line. Note that the breakdown times within each pulse were observed to be uniformly distributed. Fig.3 shows the average breakdown rate as a function of accelerating gradient after 500 and 1500 hours of processing respectively. Fig.4 shows the breakdown rate dependence on pulse length for various structures at defined practical gradients.

Other experiments at CERN for the CLIC project [23] done for 30 GHz structures indicate that higher gradients may be obtained with metals with higher melting points such as Molybdenum and Tungsten. At a pulse length of 16 ns, a gradient over 150 MV/m was reached with a Molybdenum structure. It is too early to reach any

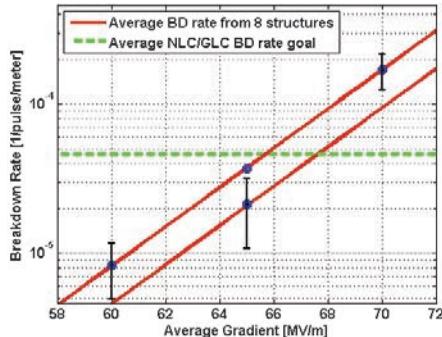


Figure 3: Average breakdown rates for a series of NLC test structures as a function of accelerating gradient after 500 hours (upper line) and 1500 hours (lower line) of rf processing.

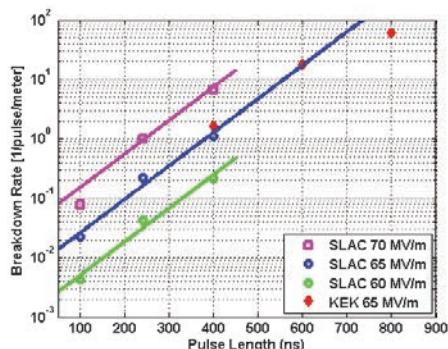


Figure 4: Breakdown rate dependence on pulse length for various NLC structures.

conclusions regarding longer pulses with these materials, nor whether the processing times will be practical.

In summary, it is expected that given enough resources, much more will be learned experimentally on the subject of rf breakdown in accelerator structures in the next few years, and that theoretical calculations combined with empirical observations may then tell us what ultimate practical gradients can be reached as a function of frequency, group velocity, power input and pulse length with any given material.

### References

- [1] R.H. Fowler and L. Nordheim, Proc. Roy. Soc., A119 (1928)
- [2] J.W. Wang, PhD Thesis, SLAC-R-339, (1989)
- [3] M. Luong et al, Proc. 7th Workshop on RF Superconductivity, (1996)
- [4] G. A. Mesyats, IEEE Trans. Electrical Insulation, EI-18 (3) (1983)
- [5] M. Jimenez et al, J. Phys. D 27 1038 (1994)
- [6] N.S. Xu, in High Voltage Vacuum Insulation, Academic Press (1995)
- [7] C.S. Athwal et al, IEEE Trans. Plasma Sci. PS-13(5) (1985)
- [8] R.V. Latham, Proc. XIIth Int. Symp. Discharges & Electrical Insulation in Vacuum (1986)
- [9] K.H. Bayliss, R.V. Latham, Proc. Roy. Soc. A403 (1986)
- [10] G.A. Loew, J.W. Wang, SLAC-Pub-4647 (1988)
- [11] A. Septier, P.M. Lapostolle, Linear Accelerators, North-Holland Pub. (1970)
- [12] S. Takeda, VIIth Int. Workshop on Linear Colliders, Zvenigorod (1997)
- [13] P. Wilson, SLAC-PUB-3637 (Rev.)
- [14] Pritzkau, PhD Thesis, SLAC-R-577 (2001)
- [15] Le Pimpec et al, SLAC-PUB-9526 (2002)
- [16] V. Dolgashev, S. Tantawi, SLAC-PUB-10175 (2003)
- [17] V. Dolgashev et al, Applied Physics Letters, Vol. 97, No.17 (2010)
- [18] J. Knobloch, PhD Thesis, Cornell U., CLNS Thesis 97-3 (1997)
- [19] P. Wilson, SLAC PUB-11086 (2005)
- [20] L. Laurent et al, PhD Thesis, UC Davis, (2002)
- [21] J.W. Wang et al, SLAC-PUB-5900 (1992)
- [22] S. Dobert, et al., SLAC-PUB-11207, (2005)
- [23] W. Wuensch et al., PAC 2003

### 6.14 HIGH VOLTAGE TECHNIQUE *B. Goddard, CERN*

#### Insulators

(i) Properties of Insulators: See Secs.6.1, 7.2.9, 8.7 for bulk properties of common dielectrics as well as [1]-[4] herein.

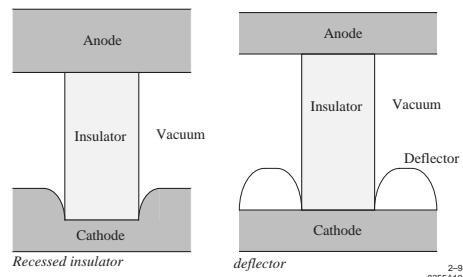
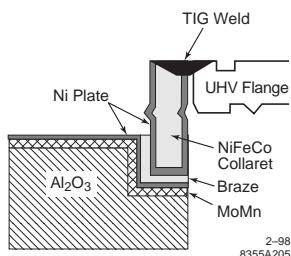
(ii) The relative dielectric strengths of gases at atmospheric pressure [4] (see also Sec.6.6) are:  $H_2 = 0.5$ ,  $N_2 = 1$ , Dry Air = 1,  $SF_6 = 2.5$ ,  $C_5F_8$  (Freon) = 5.5. (Note that by international agreement Freon is no longer tenable in new systems. See Sec.7.2.9 for alternatives.)

(iii) Alumina (97%  $Al_2O_3$  for ease of brazing) insulators are recommended for UHV applications in radiation environments ( $\alpha$ ,  $\gamma$ , e, p, n).

(iv) Design considerations for 97% Alumina (non-glazed) insulators are [5, 6]: (a) Conventional isostatic pressing and sintering preparation;

Table 1 Electrode material properties [1, 7]

| Material        | Composition %  | Density [kg/m <sup>3</sup> ] | Melting Pt. [°C] | $\kappa$ [W/m·K] | Application      |
|-----------------|--|------------------------------|------------------|------------------|------------------|
| SS, 304LN Inox  | Fe Cr <sub>18</sub> Ni <sub>12</sub> N <sub>0.16</sub> Mo <sub>2.5</sub> | 8000                         | 1280             | 16               | anode & cathode  |
| Ti UTA6V        | Ti Al <sub>6</sub> V <sub>4</sub>  | 4500                         | 1680             | 18               | anode & cathode  |
| Tungsten        | W  | 19300                        | 3370             | 160              | anode wire/foil  |
| W/Rhenium       | W Re <sub>25</sub>   | 19800                        | 3300             | 140              | anode wire/foil  |
| Peralumin,PRE30 | Al Mg <sub>3</sub> Cu <sub>0.1</sub> Si <sub>0.5</sub> Mn <sub>0.4</sub> | 2700                         | 660              | 160              | oxidized cathode |

Figure 1: Assembly of UHV flange with Al<sub>2</sub>O<sub>3</sub> insulator (deflector not shown).

(b) Surface finish obtained by grinding followed by sandblasting with Al<sub>2</sub>O<sub>3</sub> powder; (c) Ceramic-metal junctions made by brazing to Ni plated, Mo-Mn metallisation on ceramic - Fig.1; (d) Reduce field at Cathode/vacuum/insulator triple junction via deflectors or recess - Fig.2; (e) Avoid all organic and metallic contamination at all stages: final step UHV bakeout at least 350°C.

### High voltage electrodes

(i) Bulk properties of common electrode materials are shown in Tab.1.

(ii) Design considerations for high voltage electrodes [7]-[11]: (a) Mechanical polishing should not use any organic material in paste. Main surface features can be removed using ScotchBrite<sup>TM</sup>; (b) Mirror finish (electropolished) is not necessary for large devices; (c) Avoid organic contamination at all stages; (d) Final step UHV bakeout at least 300°C.

(iii) Anodic oxidation of aluminium cathodes can provide increased high voltage holdoff capability - Fig.3 [7]-[10].

#### Processes for anodic oxidation [7]-[10]

(i) small gap ( $d < 1\text{-}2$  cm), chromic acid bath (140 g/l) at 38 °C, seal off in pure boiling water for 1 hour; (ii) large gap ( $d > 1\text{-}2$  cm), sulphuric acid bath (20% b.w.) at 20°C, seal off in pure boiling water for 1 hour.

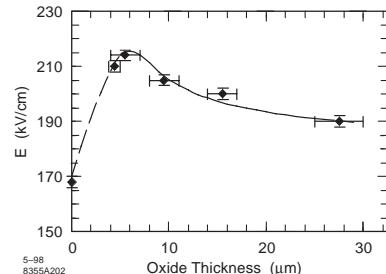


Figure 3: Conditioned breakdown field across 3 cm gap as a function of thickness of cathode alumina coating.

**High voltage cables** Capacity per meter  $C$  of coaxial cable (conductor and screen diameters  $r$  and  $R$ , dielectric  $\epsilon_r$ ):

$$C = \frac{2\pi\epsilon_0\epsilon_r}{\ln(R/r)} \quad (1)$$

**High voltage generators** Output of  $n$  stage Cockcroft-Walton multiplier (unit capacity  $C$ , supply frequency  $f$ , input voltage  $V$ , load  $I_0$ ) [13]:

$$V = 2nV - \frac{I_0}{fC} \left( \frac{2n^3}{3} + \frac{n^2}{2} + \frac{n}{6} \right) \quad (2)$$

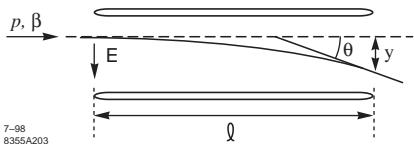


Figure 4: Deflection in an electric field.

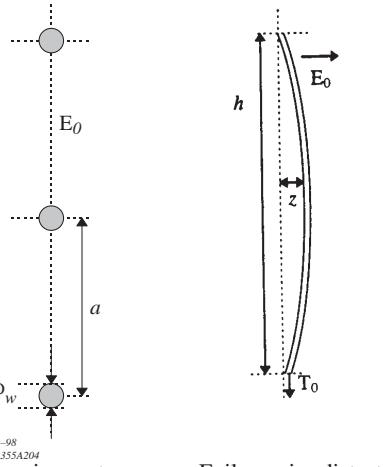


Figure 5: (left) Thin wire septum, (right) Distortion of thin foil or wire septum in electric field.

Optimum number of stages  $n_{\text{opt}}$  for nominal output  $V_{\text{nom}}$  to minimize output ripple (assuming  $n^3 \gg n^2$ ) [13]:  $n_{\text{opt}} = \sqrt{V_{\text{nom}} f C / I_0}$ .

**High voltage high field devices** (i) Deflection  $\theta, y$  of a charged particle in electrostatic field  $E_0$  - Fig.4:

$$\theta = \tan^{-1}(E_0 l / p\beta); \quad y = E_0 l^2 / 2p\beta \quad (3)$$

(ii) Effective potential  $V_{\text{eff}}$  of stretched septum wire due to penetration of gap field  $\phi_w \ll a$ ,  $h \gg a$  - Fig.5 [8]:

$$V_{\text{eff}} \approx -\frac{E_0 a}{2\pi} \ln \frac{\pi \phi_w}{a} \quad (4)$$

(iii) Distortion  $z$  of thin septum foil tensioned at  $T_0 [\text{Nm}^{-1}]$  - Fig.5 [8]:

$$z = \epsilon_0 E_0^2 h^2 / 16 T_0 \quad (5)$$

(iv) Distortion  $z$  of stretched septum wire diameter  $\phi_w$ , spacing  $a$ , tensioned at  $T_0 [\text{Nm}^{-1}]$  [8].

$$z = \epsilon_0 a E_0^2 h^2 / 4\pi \phi_w T_0 \quad (6)$$

(v) Electrostatic energy  $W_w$  per metre of wire septum [8]:

$$W_w = C_w V_0^2 / 2 \approx \epsilon_0 a V_0^2 / 2h \quad (7)$$

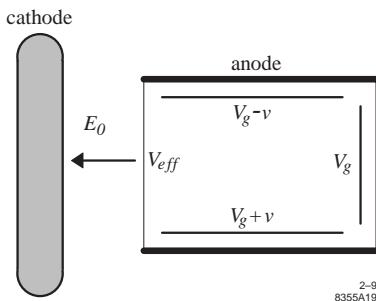


Figure 6: Ion trap potentials.

(vi) Ion Trap electrode potentials - Fig.6 [8].

$$V_g = V_{\text{eff}}, \quad v \approx 0.1 V_g \quad (8)$$

(alternate sign of  $v$  for up and down plates if several devices are used).

#### High voltage high field device performance

(i) Fowler-Nordheim field emission from metal surface with work function (valid for  $T < 600\text{K}$ ,  $j < 10^{12} \text{ A/m}^2$ ) [see also Eq.(1), Sec.6.13]:

$$j = 1.54 \times 10^{-6} \frac{E^2}{\phi t^2(y)} \exp\left(-6.83 \times 10^9 v(y) \frac{\phi^{\frac{3}{2}}}{E}\right) \quad (9)$$

(ii) Functions used in the Fowler-Nordheim equation [14]:

$$y = 3.79 \times 10^{-5} \sqrt{E}/\phi \quad (10)$$

$$t(y) \approx -0.0020y^3 + 0.0131y^2 + 0.1091y + 0.9899$$

$$v(y) \approx 0.0445y^3 - 0.6782y^2 - 0.4091y + 1.0437$$

(iii) Work function  $\phi$  of electrode metals:

| Metal | $\phi$ [eV] | Metal | $\phi$ [eV] |
|-------|-------------|-------|-------------|
| Au    | 4.82        | Ni    | 4.6-5.0     |
| Cr    | 4.60        | Pt    | 5.32        |
| Cu    | 4.46        | Sn    | 4.38        |
| Fe    | 3.9-4.8     | Ta    | 4.19        |
| Mg    | 3.68        | W     | 4.0-4.5     |
| Mo    | 4.2         | Zn    | 3.0-4.0     |

#### High voltage conditioning procedures

(i) Conditioned voltage  $V_{\text{con}}$  should exceed operational voltage  $V_{\text{op}}$  by 10 to 50% [4],[7]-[13] depending on the performance level required and observed deconditioning rate of the device;

(ii) Current conditioning:  $V$  increased in small steps with  $I$  allowed to stabilize (a few mA) at each step - Fig.7. A stable  $I$ - $V$  characteristic will result for  $V < V_{\text{con}}$  [10]-[13];

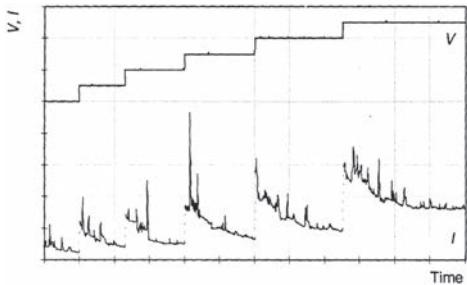


Figure 7: Example of current conditioning of high voltage gap for 5 m long 200 kV separator with 4 cm gap.

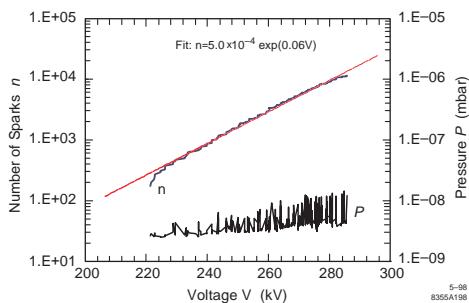


Figure 8: Spark conditioning of 15m long electrostatic septum, 20 mm gap, showing strong exponential increase in number of sparks required per kV.

(iii) Gas conditioning:  $V$  increased in small steps with  $I$  allowed to stabilize (a few mA) at each step, with residual (e.g. He, Ar) gas pressure of  $10^{-4}$  -  $10^{-5}$  mbar. Subsequent temperature  $< 100^\circ\text{C}$  to avoid deconditioning [10]-[13];

(iv) Spark conditioning: For a fixed low generator current  $I$ ,  $V$  is allowed to increase until the gap breaks down; procedure is repeated until  $V = V_{\text{con}}$ . The capacity of the external circuit should be minimized to limit the energy dissipated per spark to avoid electrode damage (e.g. using decoupling resistors); the vacuum pressure  $P$  should be monitored, preferably with a feedback to reduce  $I$  if  $P \gtrsim 10^{-8}$  -  $10^{-7}$  mbar. The voltage ( $V_{\text{con}}$ ) gain per spark decreases exponentially - Fig.8 [4], [10]-[13].

### Basic performance

(i) Holdoff voltage of uniform field gap insulated with SF<sub>6</sub> or N<sub>2</sub> increases linearly with pressure - Fig.9 [4];

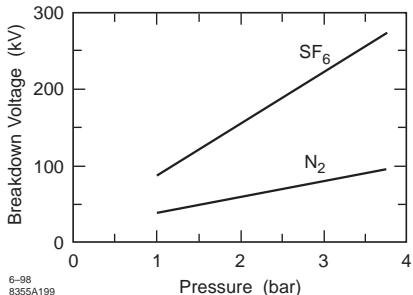


Figure 9: Breakdown in SF<sub>6</sub> and N<sub>2</sub> vs pressure.

(ii) Holdoff voltage  $V$  as approximate function of uniform field gap distance  $d$  [4, 7, 12, 13]:  
 (a)  $V \propto d$ ;  $d \leq 1$  cm in vacuum;  
 (b)  $V \propto \sqrt{d}$ ;  $d \geq 1$  cm in vacuum;  
 (c)  $V \approx 24.4 d + 7.5\sqrt{d}$ ;  $d < 16$  cm in dry air at 1 atm. 25°C,  $V$  in kV,  $d$  in cm.

(iii) Holdoff voltage  $V_S$  as a function of area  $S$  [7]:  $V_{S_2}/V_{S_1} = (S_2/S_1)^\mu$  for  $1 < S < 1000$  cm<sup>2</sup>,  $1 < d < 10$  cm. ( $\mu \approx 0.05$  for Inox anode, oxidized peralumin cathode,  $d \sim 1$  cm;  $\mu \approx 0.1$  for Inox anode and cathode).

(iv) For rate of ionization by beam see Secs.3.3.8, 3.3.9.

### Performance in accelerators (Sec.7.2.9)

(i) For unipolar devices exposed to (synchrotron radiation) photons and photoelectrons, spark rates with negative high voltage can be  $10^4$  times higher than with unipolar positive or both negative and positive high voltage [15];

(ii) Spark rate increases approximately exponentially with field, with and without beam. For a given inter-electrode field, a large gap with a

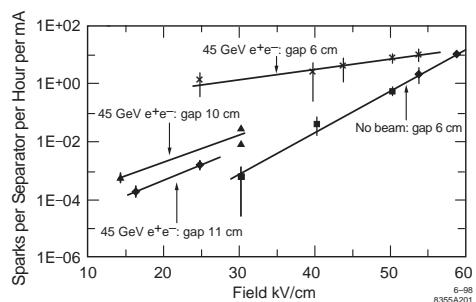


Figure 10: Performance of vertical electrostatic separators, 4 m length, in LEP for varying interelectrode gap and field.

higher voltage can give better results in the accelerator environment than a small gap with lower voltage - Fig.10 [15];

(iii) Oxidized cathodes are extremely sensitive to direct synchrotron radiation photons [16]. Cathodes must be protected by equipment layout or mechanical shielding.

## References

- [1] E.R. Cohen, The Physics Quick Reference Guide, AIP (1996)
- [2] Reference Data for Radio Engineers, ITTC, Howard Sams & Co. INC. (1975)
- [3] R.W. Silars, Electrical Insulating Materials and Their Application, P.Peregrinus/IEE (1973)
- [4] T.J. Gallagher, A.J. Pearmain, High voltage: Measurement, Testing and Design, Wiley (1983)
- [5] N. Garrel et al, CERN SL/Tech. Spec. 94-02 (BT) (1994)
- [6] S. Whitehead, Dielectric Breakdown of Solids, Clarendon (1953)
- [7] F. Rohrbach, Isolation Sous Vide CERN 71-05 (1973)
- [8] M. Thivent, CERN PS/PSR/Note 83-08 (1983)
- [9] F. Rohrbach, CERN 64-50 (1964)
- [10] L.L. Alston (ed), High Voltage Technology, Oxford (1968)
- [11] R.V. Latham High Voltage Vacuum Insulation: The Physical Basis, Academic (1981)
- [12] R.V. Latham (ed), High Voltage Vacuum Insulation, Basic Concepts and Technological Practice, Academic (1995)
- [13] E. Kufeland, M. Abdullah, High Voltage Engineering, Pergamon (1970)
- [14] Values of Fowler-Nordheim field emission functions  $v(y)$ ,  $t(y)$  and  $s(y)$ , General Electric Technical Information Series, Report 66.C-148 (1966)
- [15] N. Garrel et al, Le Vide: Science, Tech. et Appl., 275 (1995) p.386
- [16] W. Kalbreier, B. Goddard, IEEE Trans. Elec. Insul. 28 (4) (1993) 444

## 6.15 COATING RECIPES

### 6.15.1 Recipes for Coating Windows

*R.M. Sundelin, H.L. Phillips, TJNAF*

Windows used for transmitting rf or microwave power and for isolating vacuum from atmosphere are normally made of a high-alumina ceramic, but are sometimes made of beryllium oxide, aluminum nitride, sapphire or quartz. These

ceramics have high secondary emission coefficients, enhanced for tangential impact, and tend to multipactor (Sec.6.9). The multipactor produces local heating, which can crack the ceramic due to thermal stresses. The multipactor can also initiate gas discharges which, if they rf-sputter metal from locations near the window, can deposit enough metal on the window to cause excessive rf heating, also leading to ceramic cracking.

The coating itself can be a serious and damaging source of heat if metallic in nature and coating thickness will be critical. For TiN there is an optimum thickness in the tradeoff between the secondary electron yield (SEY) coefficient and window heating [1] but other coatings exist as oxides with similar SEY coefficients [2].

Ti N is often assumed to provide charge drainage which is rarely true unless quite thick [3]. A good alternative is a highly controllable surface resistance through ion implantation [4] coupled with a low SEY oxide.

Coatings designed to have secondary emission coefficients less than 1 can suppress multipacting (Sec.6.9). Several materials and methods of coating deposit have been employed for this purpose. Below they are listed with their advantages (+) and disadvantages (-).

Titanium: (+) effectively suppresses multipacting with low loss at proper thickness, (-) thickness must be accurately controlled, oxidizes when exposed to air and must be deoxidized by baking at  $> 600^{\circ}\text{C}$  or by rf processing. Rf processing requires care to avoid arcing damage to the layer or cracking of ceramic.

*Recipe 1:* Deposit 100 Å Ti [5].

*Recipe 2:* DC sputter in  $5 \times 10^{-2}$  Torr Ar until  $1.6 \times 10^8 \Omega/\text{sq}$  (DC) is achieved [6].

$\text{Ti}_x\text{O}_y$ : (+) thickness less critical if not baked above  $500^{\circ}\text{C}$ . Rf reduction to Ti appears to be self limiting when multipactor stops (advantage not realized if  $> 600^{\circ}\text{C}$  bake carried out); (-) Oxidizes when exposed to air. Must be deoxidized by  $> 600^{\circ}\text{C}$  bake or rf processing. The latter requires care to avoid arcing damage to layer or ceramic cracking.

*Recipe:* Reactively sputter Ti using crossed field diode sputtering array, 3.5 kV, 0.06 T,  $5 \times 10^{-5}$  Torr, 90% O<sub>2</sub> 10% Ar. 100-200 Å [7].

$\text{Ti}_x\text{N}_y$ : (+) High thermal stability with minimal processing required; (-) Stoichiometry not well defined - reactive sputtering yields different results than sputtering TiN and thickness must be

accurately controlled to avoid excessive rf heating.

Recipe: Reactively sputter Ti using crossed field diode sputtering array, 3.5 kV, 0.06 T, 5 × 10<sup>-5</sup> Torr, 90% N<sub>2</sub>, 10% Ar. 80-100 Å[7].

Cr<sub>2</sub>O<sub>3</sub>: (+) Stable against high temperatures with low loss tangent; (-) Earlier results [7] not as good as more recent ones [8].

Recipe: Post-oxidize 30 Å Cr film.

Gold black: (+) low additional loss; (-) potential source of particles - tested only for high average power, not high peak power.

Recipe: Evaporate gold 15 cm from window in 0.5 Torr of N<sub>2</sub>. Deposit a layer several 10's of angstroms thick. Remove loose particles with gas jet [9].

Microgrooving: (+) Can be used in combination with other techniques to reduce thickness (and associated dissipation) of other coatings (see Sec.6.9); (-) Insufficient effective secondary emission coefficient by itself. Coating walls of grooves increases coating complexity. Ridges cause local field enhancement thereby reducing utility for high peak power.

Recipe: Grind grooves into ceramic at right angles to electric field projection onto ceramic surface. Period of grooves should be short compared to multipacting trajectory length.

## References

- [1] S. Michizono, et al., "Ti N Film Coating on Aluminum Radio Frequency Windows", J. Vac. Sci. Tech. A 10(4) 1992, 1180
- [2] A.R. Nyairesh, E.L. Garwin, F.K. King, and R.E. Kirby, "Properties of Thin Antimultipactor Coatings for Klystron Windows", SLAC-PUB-3760, 1985
- [3] Y. Saito, et al., 15th Int. "Symposium on Discharges and Electrical Insulation in Vacuum", Darmstadt, 1992
- [4] S. Anders, et al., "Surface Resistivity Tailoring of Ceramic Accelerator Components," Proc. of the 1993 Particle Acc. Conf., 1390.
- [5] Bud Hjelm, Litton Industries, private communication; K. Primdahl et al, PAC 95, p.1687
- [6] George Dôme, CERN, private communication
- [7] K. Welch, SLAC-174 (1974)
- [8] A. Nyairesh et al, J. Vac. Sci. & Tech. (1986); E. Garwin, SLAC, private communication
- [9] Godfrey Saxon, Daresbury Lab.

## 6.15.2 Recipes for Coating Ceramic and Metal Vacuum Chambers

S.D. Henderson, FNAL

A thin-film coating may be applied to the inner surface of a vacuum chamber for the following purposes: (i) to provide a conducting surface to carry the beam image current in the case of ceramic vacuum chambers, (ii) to absorb scattered x-rays in the case of central beampipes for experimental detectors at colliders, and (iii) to exploit favorable vacuum properties (reduced outgassing, reduced photon induced desorption, or gettering) or electrical properties (lower resistivity or secondary electron emission) of the coating material.

The vacuum requirements of a coated chamber are the same as those of any other UHV system component (Sec.5.10) with one additional requirement: adhesion of the film is extremely important to avoid the generation of small particles in the accelerator vacuum system.

### Choice of material and thickness

Ceramic Vacuum Chambers Ceramic chambers are used for pulsed/ac magnets, feedback kickers, shakers, etc. to allow external time-varying fields to penetrate the vacuum chamber. A thin metallic coating is required to carry the beam image current (for beam stability) and to protect external components from the beam fields. See also Sec.3.2.4.

The coating material and thickness, *d*, are chosen based on the following requirements: (i) external fields must penetrate the coating, (ii) the risetime of the external field must be preserved (for fast kickers), (iii) electrical continuity to the adjoining vacuum chambers must be provided, (iv) the beam fields must be contained, (v) the film must tolerate magnet pulsing (remain adherent – no arcing) subject to acceptable heat loads from, (vi) ohmic heating from the beam image current, (vii) eddy current heating from magnetic pulsing, and (viii) heating from absorption of HOM power generated by discontinuities at the joints between the chamber and the surrounding vacuum system. It should be emphasized that the beam impedance of the coated ceramic chamber as a whole must be considered to evaluate the HOM losses and possible resonance effects [1, 2].

The following are useful for evaluating material and thickness selection: 1. Skin Depth:

$$\delta = \sqrt{\frac{2\rho_r}{\omega\mu}} = 1.59\mu\text{m} \sqrt{\frac{\rho_r[\mu\Omega\text{ cm}]}{(\mu/\mu_0)f[\text{GHz}]}} \quad (1)$$

$$\delta_B = 0.728\mu\text{m} \sqrt{\frac{\rho_r[\mu\Omega\text{ cm}]\sigma_z[\text{cm}]}{(\mu/\mu_0)}} \quad (2)$$

(characteristic bunch frequency)

## 2. Sheet Resistance:

$$R_s = \rho_r/d \quad (3)$$

3. Risetime of Magnetic Fields Inside Chamber due to Eddy Current Effect [1]:

$$\tau_B = \mu_0 r / 2R_s = 6.28 \text{ ns} \quad r[\text{cm}]/R_s[\Omega/\text{sq.}]$$

(circular pipe - step pulse) (4)

4. Penetration of beam fields (Piwinski's criterion) [3]: Beam field penetrates the coated ceramic chamber when  $d < \delta_B^2/d_c \equiv d_P$  (see also [2]), where  $d_c$  is the ceramic thickness.

5. Ohmic heating of a circular pipe of radius  $r$  (three regions distinguished):

(a)  $d \gg \delta_B$  (pure metallic pipe) [3, 2]:

$$\begin{aligned} \frac{P}{l} &= \frac{n_b I_0^2 C}{4\pi^2 r \sigma_z^{3/2}} \sqrt{\frac{Z_0(\mu/\mu_0)\rho_r}{2}} \Gamma\left(\frac{3}{4}\right) \quad (5) \\ &= \frac{4.26 n_b I_0 [\text{A}]^2 C [\text{m}]}{r [\text{cm}] \sigma_z [\text{cm}]^{3/2}} \sqrt{\frac{\mu}{\mu_0} \rho_r [\mu\Omega\text{ cm}]} \left[\frac{\text{W}}{\text{m}}\right] \end{aligned}$$

(b)  $d_P < d \lesssim \delta_B$ :

$$\begin{aligned} P/l &= \frac{n_b I_0^2 C}{2\sqrt{\pi}\sigma_z} \frac{R_s}{2\pi r} \\ &= 449 \frac{n_b I_0 [\text{A}]^2 C [\text{m}]}{r [\text{cm}] \sigma_z [\text{cm}]} R_s [\Omega/\text{sq.}] \left[\frac{\text{W}}{\text{m}}\right] \quad (6) \end{aligned}$$

(c)  $d < d_P$  [3]: Films made using the technology described above will be continuous and sufficiently rugged only for  $d > 100 \text{ \AA}$ . As a result, this case is of little practical concern.

6. Eddy Current Heating (pulsing at repetition frequency  $f$ ) – two waveforms considered, half sine (frequency  $\omega$ ) and square pulse (risetime  $\tau$ ) ( $B_0 = \max B$  field).

Instantaneous power dissipated in a resistive loop of length  $l$ , width  $w$ , height  $h$  (MKS units):

$$P = \frac{\dot{B}^2 w^2 l h}{2R_s} \quad (7)$$

Time-averaged power per unit length for a circular pipe (radius  $r$ ):

$$P/l = f B_0^2 \frac{\pi r^3}{R_s} \begin{cases} \frac{\pi\omega}{2} \text{ half-sine} \\ \frac{2}{\tau} \text{ square pulse} \end{cases} \quad (8)$$

and rectangular pipe ( $w \times h$ ):

$$P/l = f B_0^2 \frac{(3h+w)w^2}{6R_s} \begin{cases} \frac{\pi\omega}{2} \text{ half-sine} \\ \frac{2}{\tau} \text{ square pulse} \end{cases} \quad (9)$$

Minimum power density dissipation occurs when Ohmic losses equal Eddy current, however, this may not be optimal from risetime considerations. Elliptical chambers may have thinner coating on the sides to reduce eddy currents, with little impact on Ohmic heating due to geometry. Particular care should be taken in the choice of material regarding adhesion, and response of the film to the voltage pulse induced during magnet pulsing.

Ceramic chambers in use at CESR have Kovar coatings with resistivity (magnetic field risetime)  $1.3 \Omega/\text{sq.}$  (40 ns),  $0.4 \Omega/\text{sq.}$  (250 ns, 7  $\mu\text{s}$ ). Ceramic chambers in use for the LEP injection system [28] have Ti coatings with  $0.25 \Omega/\text{sq.}$  (3  $\mu\text{s}$ ). Ceramic chambers in use at APS [21] were coated with commercially available resistive paste and have resistivities  $50 \Omega/\text{sq.}$  (100 ns),  $25 \Omega/\text{sq.}$  (150 ns), and  $0.1 \Omega/\text{sq.}$  (1  $\mu\text{s}$ ).

Requirements for interaction region beam-pipes Central beampipes of  $e^+e^-$  collider detectors are often made of beryllium and may have a metallic coating applied to the inner surface for the absorption of scattered synchrotron radiation (Sec.3.3.9). The requirements for a coated detector beampipe are as follows:

(i) Total thickness in radiation lengths of beampipe and x-ray absorbing material is minimized. A “budget” is constructed based on detector charged particle tracking requirements and detector background occupancy and radiation dose goals.

(ii) The coating material is chosen which maximizes the absorption of scattered x-rays for the minimum radiation length. The beampipe coating material must be a good match to scattered fluorescent photons from surrounding materials.

(iii) The coating material must be a good conductor to minimize Ohmic heating.

(iv) Detector tracking requires uniform coating thickness ( $\lesssim 20\%$ ).

The central Be beampipe for the CLEO detector at CESR has a  $10 \mu\text{m}$  Au coating ( $0.3\% X_0$ ), as will the BaBar beampipe at PEP-II. A previous CLEO beampipe used  $25 \mu\text{m}$  Ag.

**Specialized vacuum chamber coatings** There is active interest in identifying coatings which offer superior vacuum properties over uncoated aluminum or stainless steel. In particular, low photo-stimulated desorption coatings are under study [4] with potential application for high current storage rings and light sources. TiN films deposited on stainless steel have shown lower outgassing rates [5]. The PEP-II low energy ring and the SNS accumulator ring chambers are coated with TiN to reduce secondary electron emission [6, 7]. High RRR Cu films have been electrodeposited on beam chambers for HERA [11]. Non-evaporable getter coatings of TiZrV and related compounds have low activation temperatures and low secondary electron yields [8]. TiZrV coatings have been applied to insertion device beampipes [9], and are being used at the LHC [10].

**Coating options** Thin-films may be deposited either through physical deposition methods (sputtering [12, 13, 14], evaporation [15]), chemical methods (electrodeposition, chemical vapor deposition [16], sol-gel [17]), or hybrid processes. For an overview of deposition methods and experimental techniques, see [18, 19, 20].

Sputtering is a natural choice for depositing films on the inside of a vacuum chamber for the following reasons: i) the cylindrical or oval chamber geometry is accommodated by sputtering a target rod mounted along the axis of the chamber, or by using a sliding sputtering gun, ii) sputtered coatings *if properly deposited* are extremely adherent, have high purity and low outgassing rates, iii) deposition rates can be quite high, and iv) high temperatures can be avoided, unlike many of the other processes (important for coating assembled chambers which include welds and braze joints).

Other techniques can and have been used, however. The recent development of resistive pastes is an attractive alternative for ceramic coatings [21]. The emphasis here, however, is on sputter deposition, since this is a common “in-house” method of preparing high quality films.

**General sputtering considerations** [12, 22] There are 3 varieties of sputter deposition: DC diode, Magnetron and RF, as well as several hybrid processes. All the processes mentioned rely on a glow discharge established between a cath-

ode of the coating material and the vacuum chamber. DC diode and magnetron sputtering are used for depositing coatings of pure metals and alloys, and a number of compounds. RF sputtering is used for depositing insulating film coatings. Thin-films of compounds (Nitrides, Oxides, Carbides) may be produced by sputtering a metal target in the presence of a reactive gas, such as Ti sputtering in Ar/N<sub>2</sub> for TiN films.

Given a cathode target material and sputtering gas (assumed to be Ar), the parameters which influence the glow discharge behaviour are the gas pressure ( $p$ ), discharge voltage ( $V$ ), cathode current ( $I$ ), anode-cathode separation ( $d_{ac}$ ) and magnetic field strength ( $B$ ).

In a DC diode discharge in Ar, the dark space thickness,  $d$ , is related to pressure by  $pd \simeq 300 \text{ mTorr cm}$ . The pressure required to sustain a discharge in a tube of radius  $r$  is  $p \gtrsim 300/r[\text{cm}] \text{ mTorr}$ . Most sputtering systems operate to the left of minimum of the Paschen curve, so high voltages are required to sustain a discharge  $V \simeq 2 \text{ kV}$ . Maximum sputtering rates occur near  $p \simeq 50 \text{ mTorr}$ .

In a Magnetron Discharge [22]:

1. Gyro radius of secondary electrons emitted from cathode:  $r_g[\text{cm}] = 3.37\sqrt{V[\text{V}]/B[\text{G}]}$ .

2. Turning point distance of secondary electron orbit:  $d_t = r_g$  (planar magnetron or cylindrical with  $R \gg d_t$ )  $d_t = 2r_g$  (cylindrical magnetron with  $R \ll d_t$ ).

3. Minimum magnetic field required for sustained cylindrical magnetron discharge in pipe of radius  $r$ :  $B[\text{G}] \gtrsim 10\sqrt{V[\text{V}]/r[\text{cm}]}$ .

4. The minimum voltage for a sustained discharge at typical pressures in Ar is about 600 Volt.

5. Current-voltage characteristic:  $I \propto V^n$ ,  $5 \lesssim n \lesssim 9$ .

The operating conditions which affect film structure, adhesion, and gas incorporation are: (i) deposition rate, (ii) gas pressure, (iii) substrate temperature, (iv) substrate bias voltage, and (v) vacuum system impurity concentration. Particular care must be taken to minimize the introduction of impurities during the coating process. UHP grade Ar (99.999%) or better should be flowed at high throughput during the deposition process. Substrate surface cleanliness prior to deposition may be the most important factor in film adhesion. Substrate surface preparation should follow standard UHV techniques. The vacuum system should be baked at  $110^\circ\text{C}$  prior to deposition.

Metal substrates should be “backspattered” (ion-etched) and targets should be “pre-sputtered” to remove surface impurities before deposition.

**Thin-film characterization techniques** Properties of thin-films are reviewed in [23, 24, 25] Tab.1 summarizes common thin-film characterization methods [23, 20]. Many of the tests cannot be performed on the inside of a coated chamber, but are useful for evaluating prototype coatings. Tests may be applied to witness coupons (identical to the bulk material) which have been coated during the final production coating.

### Coating recipes

Kovar coated ceramic chamber (CESR) [26]  
 $(R_s = 0.4\Omega/\text{sq.})$  Method: DC Diode sputtering of Kovar in Ar. Kovar rod (cathode, 0.375" dia.) and 2 copper anode wires (0.020" dia.) mounted coaxially in chamber.

Chamber dimensions: Oval,  $9 \times 5 \times 60\text{cm}$

Procedure:  $110^\circ\text{C}$  vacuum bake 5 hrs. Cool.

$P_{Ar} = 75 \text{ mTorr}$ ,  $V = 2.0 \text{ kV}$ ,  $I = 55 \text{ mA}$  ( $0.25 \text{ mA/cm}^2$ )

Sputter at  $0.06 \mu\text{m}/\text{hr}$  for 20 hours

Vacuum bake at  $110^\circ\text{C}$ , cool, backfill  $N_2$ .

Ti coated ceramic chambers (LEP)[27, 28]  
 $(R_s = 0.25\Omega/\text{sq.})$  Method: Magnetron sputtering in Ar. Custom water cooled magnetron gun using SmCo magnet and Ti target in shape of chamber cross-section.

Chamber dimensions:  $20 \times 5 \times 130 \text{ cm}$ .

Procedure: Assembly vacuum baked and maintained at  $110^\circ\text{C}$  during deposition.

$P_{Ar} = 1.5 \text{ mTorr}$ ,  $V = 400 \text{ V}$ ,  $I = 2.5 \text{ A}$  ( $\sim 35 \text{ mA/cm}^2$ ).  $B = 700 \text{ G}$  at target surface. Ar flow 80 cc/min. Gun moved at 3 mm/min.

Resistive paste coated ceramic (APS)[21]

$(R_s = 0.1 - 50 \Omega/\text{sq.})$  Method: Brush application of resistive paste (manufactured by Heraeus Cermalloy).

Procedure: Thin paste by 30% with Terpineol. Paste is pumped through nozzle to a brush which is pulled through tube, pushed back, pulled through and out. Chamber is baked at  $150^\circ\text{C}$  for 15 minutes and then oven fired as follows ( $\rightarrow$  indicates transition, with transition time above):

$25^\circ\text{C} \xrightarrow{4\text{hr}} 250^\circ\text{C} \xrightarrow{1\text{hr}} 250^\circ\text{C} \xrightarrow{6.5\text{hr}} 650^\circ\text{C} \xrightarrow{2\text{hr}}$   
 $650^\circ\text{C} \xrightarrow{15\text{hr}} 25^\circ\text{C}$

The first step in the cycle is performed in 5 Torr  $O_2$  (1 l/min flow), and the remaining steps are in ambient air. Note: this firing cycle differs from the manufacturer's.

### Gold coated beryllium beampipe (CESR)

[35, 36] ( $5 \mu\text{m}$  Au) Method: Cylindrical magnetron sputtering of axially mounted water-cooled Au rod (1 mm thick 99.999% Au clad on 0.25 in. stainless steel tube; supplied by Noble-Met Ltd., Roanoke VA). Cathode is mounted inside electrically isolated Be beampipe, which is axially mounted in a chamber inside a solenoid. Thin Cr layer is applied by sputtering (99.99% 0.25 in. dia. Rod; supplied by Angstrom Sciences, Pittsburgh, PA) to improve adhesion of Au to Be.

Procedure: i) Cr rod mounted in Be beampipe, pumped down, system flushed with Ar, ii) back-sputter Be beampipe for 15 minutes:  $B = 0 \text{ G}$ ,  $P_{Ar} = 275 \text{ mTorr}$ ,  $V = 250 \text{ V}$ ,  $I = 0.4 \text{ A}$ , iii) Apply 150 nm Cr:  $B = 250 \text{ G}$ ,  $P_{Ar} = 4 \text{ mTorr}$ ,  $V = 600 \text{ V}$ ,  $I = 0.17 \text{ A}$  for 9 minutes ( $1.7 \text{ mA/cm}^2$ ) deposition rate  $< 1 \mu\text{m/hr}$ , iv) replace Cr rod with Au rod, pumpdown and flush, v) Backsputter beampipe for 15 min:  $B = 0 \text{ G}$ ,  $P_{Ar} = 275 \text{ mTorr}$ ,  $V = 250 \text{ V}$ ,  $I = 0.4 \text{ A}$ , vi) Apply  $5 \mu\text{m}$  Au:  $B = 250 \text{ G}$ ,  $P_{Ar} = 4 \text{ mTorr}$ ,  $V = 750 \text{ V}$ ,  $I = 0.20 \text{ A}$  for 90 minutes ( $1.4 \text{ mA/cm}^2$ ) deposition rate  $< 3.3 \mu\text{m/hr}$ , vii) Backfill with Ar, cool, viii) Bake coated assembly in vacuum furnace at  $300^\circ\text{C}$  for 1 hr.

### TiN coated stainless steel vacuum chambers

[7] (100-200 nm TiN) Method: Reactive sputtering of Ti in cylindrical magnetron geometry for coating a 4 m long, 8 in. dia stainless steel chamber. A water-cooled cathode was constructed from  $1\frac{1}{2}$  in ID Ti tubing, into which Alnico permanent magnets ( $1\frac{1}{4}$  in OD, 2 in length, spaced by  $\frac{1}{2}$  inch) were inserted, producing a field of several hundred Gauss at the cathode surface.  $N_2$  was introduced from a Ti tube running the length of the cathode with holes spaced every 6 in.  $N_2$  flow controlled by partial pressure feedback, using mass flow controller and quadrupole RGA. Film has Ti/N ratio 1:1

Procedure: i) chamber is vacuum baked at  $250^\circ\text{C}$  for  $> 24$  hrs, ii) glow discharge clean for 5 minutes:  $I = 1 \text{ A}$ ,  $V = 324 \text{ V}$ ,  $P_{Ar} = 23 \text{ mTorr}$ , Ar flow = 53 sccm, iii) apply TiN,  $I = 10 \text{ A}$ ,  $V = 304 \text{ V}$ ,  $P_{tot} = 5 \text{ mTorr}$ , Ar flow = 16 sccm,  $N_2$  flow = 5.6 sccm, deposition rate  $0.1 \mu\text{m/hr}$ .

### TiZrV coated stainless steel vacuum chambers

[9, 37] ( $2 \mu\text{m}$ ) Method: DC cylindrical magnetron sputtering. The cathode consists of three twisted 1-mm dia. Ti, Zr and V wires (99.8% purity).

Table 1: Thin-film Characterization Methods.

| Film Property               | Technique(s)  | Comments  |
|-----------------------------|---|---|
| <b>Film Thickness [29]</b>  | Weight Measurement<br>Step Stylus<br>x-ray Backscattering [30]<br><br>x-ray Absorption [30] | Average thickness of thick films. Requires knowledge of density.<br>Require flat substrate and a step<br>Useful for coating on thick metal substrate. Measure flux of K x-rays from the substrate. Comparison of flux for coated and uncoated sections gives coating thickness.<br>For Be chambers, or other x-ray “transparent” substrates, measure attenuation of x-ray beam. |
| <b>Chemical Composition</b> | SEM/EDX<br>AES<br>XPS<br>RBS<br>SIMS  | 1 $\mu\text{m}$ depth, $Z \geq 11$ 0.1 at. % sensitivity<br>surface, $Z \geq 3$ , 0.1 at. %<br>surface, $Z \geq 3$ , 0.1 at. %<br>surface, $Z \geq 2$ , 0.1 at. %<br>surface, $Z \geq 1$ , $10^{-4}$ at. %  |
| <b>Adhesion</b><br>[31, 32] | Peel Test<br>Scribe Test<br><br>Bend Test<br>Quench test                                    | “Scotch-Tape Test”<br>Scribe parallel lines in a grid pattern using a sharp scribe and cutting to the substrate. Observe for flaking and perform peel test.<br>Bend the substrate 90°, inspect for cracks, tears, flaking<br>Repeated heating / LN <sub>2</sub> dunk cycles   |
| <b>Resistivity</b>          | Four-point probe  | $R_s = \frac{\pi}{\ln 2} \frac{V}{I}$ for correction factors [33]   |
| <b>Surface Finish</b>       | Step Stylus<br>AFM  |   |
| <b>Surface Structure</b>    | SEM   |   |
| <b>Grain Size [34]</b>      | SEM   |   |

Procedure:  $P_{Ar} = 2\text{-}7$  mTorr,  $V = 500$  V,  $B = 150\text{-}200$  G, deposition rate 0.1-0.3  $\mu\text{m}/\text{hr}$  on 60-100 mm dia pipes.

## References

- [1] S.S. Kurennoy, PAC 93, p.3420
- [2] X.E. Lin, SLAC-PUB-7924, submitted to IEEE MTT (1998)
- [3] A. Piwinski, PAC 77, p.1364
- [4] C.L. Foerster, G. Korn, AIP Proc. 236, Vacuum Design of Synchrotron Light Sources (1991)
- [5] M. Sato et al, Vacuum 47 (1996) 753
- [6] K. Kennedy et al, PAC 97
- [7] R. Todd, H. Hseuh, private communication; P. He et al, PAC 01, p.2159; H. Hseuh et al, Proc. 31st Adv. ICFA Beam Dynamics Workshop on Electron Cloud Effects, p.119 (CERN-2005-001)
- [8] C. Benvenuti et al, Vacuum 60 (2001) 57; C. Benvenuti et al, Vacuum 53 (1999) 219; C. Benvenuti, EPAC 98, p.200
- [9] P. Chiggiato, R. Kersevan, Vacuum 60 (2001) 67
- [10] A. Rossi, ECLOUD 04, p.113, CERN-2005-001
- [11] J. Cl. Puppie, W. Saxer, Proc. XV Int. Conf. High Energy Acc. (1992) (Int. J. Mod. Phys. A Proc. Suppl., 2A, 1993)
- [12] J.L. Vossen, J.J. Cuomo, in Thin Film Processes, Academic (1978)
- [13] B.N. Chapman, Glow Discharge Processes, Wiley (1980)
- [14] S. M. Rossnagel, in Thin Film Processes II, Academic (1991)
- [15] C.V. Deshpandey, R.F. Bunshah, in Thin Film Processes, Academic (1991)
- [16] K.F. Jensen, W. Kern, in Thin Film Processes II, Academic (1991)
- [17] L.C. Klein, in Thin Film Processes II, Academic (1991)
- [18] Thin Film Processes, Academic (1978)
- [19] Thin Film Processes II, Academic (1991)
- [20] L.I. Maissel, R. Glang, Handbook of Thin-Film Technology, McGraw-Hill (1970)
- [21] S. Milton, D. Warner, EPAC 94, p.2494
- [22] J.A. Thornton, A.S. Penfold, in Thin Film Processes, Academic (1978)

- [23] M. Ohring, *The Materials Science of Thin Films*, Academic (1992)
- [24] K.L. Chopra, *Thin Film Phenomena*, McGraw-Hill (1969)
- [25] L. Eckertova, *Physics of Thin Films*, Plenum (1986)
- [26] S. Henderson, Cornell CON-95-16
- [27] M. Mayer et al, EPAC 88, p.1378
- [28] G.H. Schroder et al, EPAC 88, p.1381
- [29] Standard Guide for Measuring Thickness of Metallic and Inorganic Coatings, ASTM B659-90
- [30] H.A. Liebhafsky et al, *X-ray Absorption and Emission in Analytical Chemistry*, Wiley
- [31] Standard Test Methods for Adhesion of Metallic Coatings, ASTM B571-91
- [32] B.N. Chapman, *Thin-film Adhesion*, J. Vac. Sci. Techno. 11 (1974) 106
- [33] L.B. Valdes, Proc. I.R.E. (1954) 420
- [34] Standard Test Methods for Determining Average Grain Size, ASTM E112-88
- [35] S. Henderson, S. Roberts, PAC 99, p.1351
- [36] Stuart D. Henderson, Proc. 8th Mtg. Div. Part. & Fields of APS (1994) World Scientific, p.1480
- [37] Y. Li, S. Ho, PAC 03, p.620

## 6.16 CAVITY MEASUREMENTS

*R. Rimmer, LBNL  
M. Tigner, Cornell U.*

### 6.16.1 Field Maps by Perturbation Methods

The fields in a cavity can be sampled by introducing a perturbing object along the beam pipe and observing the change in resonant frequency [1, 2, 3]. Fig.1 shows a typical set-up for the measurement.

For the case of a small non-conducting sphere, radius  $r$ , where the unperturbed field may be considered uniform over a region larger than

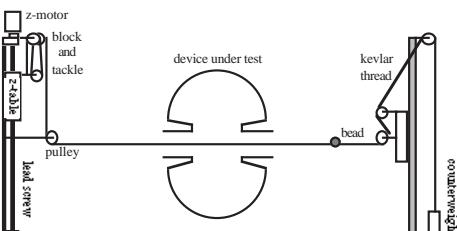


Figure 1: Bead pull apparatus for field mapping.

the bead,

$$\frac{\Delta\omega}{\omega} = \frac{\Delta U}{2U} = -\frac{\pi r^3}{U} \left( \epsilon_0 \frac{\epsilon_r - 1}{\epsilon_r + 2} E_0^2 + \mu_0 \frac{\mu_r - 1}{\mu_r + 2} H_0^2 \right) \quad (1)$$

where  $U$  is the total stored energy in the cavity and  $E_0$  and  $H_0$  are the fields at the perturber position and the perturber is so small that the fields do not vary significantly throughout its volume. For a dielectric bead ( $\mu_r = 1$ ), Eq.(1) reduces to

$$\frac{\Delta\omega}{\omega} = -\frac{\pi r^3}{U} \left( \epsilon_0 \frac{\epsilon_r - 1}{\epsilon_r + 2} E_0^2 \right) \quad (2)$$

For a metal bead ( $\epsilon_r \rightarrow \infty$ ,  $\mu_r \rightarrow 0$ ),

$$\frac{\Delta\omega}{\omega} = -\frac{\pi r^3}{U} \left( \epsilon_0 E_0^2 - \frac{\mu_0}{2} H_0^2 \right) \quad (3)$$

Shaped beads such as needles or disks can be used to enhance the perturbation and give directional selectivity [2](p.445). The enhancement or "form factor" can be calculated for ellipsoids or calibrated in a known field. Determining the direction of the field components is particularly important when dealing with the higher modes of a complex cavity shape where the measured mode frequency will not be a sure identifier of the dominant field direction.

**Shunt impedance** By mapping the longitudinal distribution of  $E_z$  and integrating along the beam path, the cavity shunt impedance can be determined,

$$\frac{R}{Q} = -\frac{\epsilon_r + 2}{\omega \pi r^3 \epsilon_0 (\epsilon_r - 1)} \frac{\left[ \int \sqrt{\frac{\Delta\omega}{\omega}(z)} e^{i k_v z} dz \right]^2}{2} \quad (4)$$

where  $v$  is the particle velocity and  $R \equiv \frac{V^2}{2P}$ , where  $V$  is the maximum energy that a particle traversing the cavity can gain, taking the finite crossing time of the cavity into account (transit time factor) and  $P$  is the power dissipated in the cavity walls to establish  $V$ . (For a cavity with uniform field and a gap length  $L$ , the transit time factor is  $T = \sin(\pi L/\lambda) / (\pi L/\lambda)$  – already taken into account in the preceding formulae.)

If the cavity is symmetric in  $z$  and  $t = 0$  at  $z = 0$  in the center, then

$$\frac{R}{Q} = -\frac{\epsilon_r + 2}{4\pi^2 f r^3 \epsilon_0 (\epsilon_r - 1)} \times \left[ \int \sqrt{\frac{\Delta f}{f}(z)} \cos\left(\frac{2\pi f z}{v}\right) dz \right]^2 \quad (5)$$

The kernel is measured at a finite number of points and numerically integrated. If the  $Q$  is then

measured, a value for  $R$ , (sometimes referred to as  $Z_{\parallel}$ ) the beam impedance, is obtained.

If the volume and dielectric constant of the perturber are not well defined, the perturber can be calibrated in a TM<sub>01</sub> cavity where the field pattern is known and  $R$  can be calculated accurately. Alternatively one can measure the integral of  $E_z^2$  through the cavity by measuring the frequency perturbation due to a thin dielectric rod passing completely through the cavity along  $z$  [4]. The small bead measurement can then be used to obtain the shape of the field for performing the time dependent integral. Such a slender dielectric rod is also very useful in determining electric field direction. An excellent material for such a rod is Sapphire with a relative dielectric constant of order 10. However, each rod must be separately calibrated in a TM<sub>01</sub> cavity as the dielectric constant is anisotropic.

These methods can easily be extended to measuring the shunt impedance of the space harmonics in periodic structures [4].

**Image effects** It is often necessary to determine field components near to a conducting surface. In this case, image effects must be taken into account [5]. For a spherical perturber with center a distance,  $d$ , from a conducting wall, the actual fields and the fields measured by frequency perturbation are related through

$$E = E_{\text{act}} \left[ 1 + \frac{1}{4} \frac{\epsilon_r - 1}{\epsilon_r + 2} \left( \frac{r}{d} \right)^3 \right] \quad (6)$$

$$H = H_{\text{act}} \left[ 1 + \frac{1}{16} \left( \frac{r}{d} \right)^3 \right] \quad (7)$$

Alternatively, one can use a hemispherical perturber drawn along the surface together with the result [6] that a hemisphere on a plane surface will produce half the frequency shift produced by a sphere of that same radius in the same field.

## References

- [1] L.C. Maier, J.C. Slater, J. App. Phys., Vol.23, No.1 (1952) 68
- [2] E.L. Ginzton, Microwave Measurements, McGraw-Hill (1957) Ch.10
- [3] W. Barry, G.R. Lambertson, PAC 87, p.1602
- [4] W.J. Gallagher, Microwave Lab. Report 767, Stanford U. (1960)
- [5] S.W. Kitchen, A.D. Schelberg, J. App. Phys., 26, 5 (1955) 618
- [6] J.C. Slater, Revs. Mod. Phys. 20 (1948) 473

## 6.16.2 $Q$ and $\beta$ Determination from Input Coupler

(See also Secs.6.6, 6.8.)

Equivalent circuit of the cavity impedance presented to the generator [1](Sec.9.3) is shown in Fig.1.

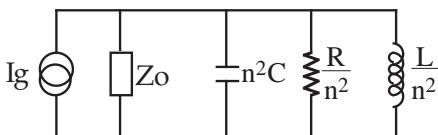


Figure 1: Cavity referred to input;  $n$  is the turns ratio of the transformer that matches the cavity to the generator with impedance  $Z_0$ .

The coupling factor is defined as

$$\beta = \frac{\text{power loss in ext. ckt.}}{\text{power loss in cavity}} = \frac{Q_0}{Q_{\text{ext}}} = \frac{R}{n^2 Z_0} \quad (1)$$

$$\frac{1}{Q_L} = \frac{1}{Q_0} + \frac{1}{Q_{\text{ext}}}; \quad Q_0 = (1 + \beta) Q_L \quad (2)$$

At resonance the cavity impedance is purely resistive (Fig.2).

$$0 < \beta < 1 \text{ (undercoupled)} \quad \text{VSWR} = 1/\beta$$

$$\beta = 1 \quad (\text{matched}) \quad \text{VSWR} = 1$$

$$\beta > 1 \quad (\text{overcoupled}) \quad \text{VSWR} = \beta$$

**Q from  $S_{11}$**  The loaded or unloaded  $Q$  can be determined from a calibrated measurement of  $S_{11}$  looking towards the cavity by finding the half power points of the reflected signal or by using the Smith chart (once the electrical delay has been adjusted to refer the impedance to the detuned-short position) [2](ch.9).

For the unloaded half-power points ( $Q_0$ ):

$$|S_{11}| = \sqrt{\frac{5S_0^2 - 2S_0 + 1}{S_0^2 - 2S_0 + 5}} \quad (3)$$

$$\text{VSWR} = \frac{2 + \beta^2 + \sqrt{4 + \beta^4}}{2\beta} \quad (4)$$

where  $S_0$  is the value at resonance. For the loaded half power points ( $Q_L$ ):

$$|S_{11}| = \sqrt{\frac{S_0 + 1}{2}} \quad (5)$$

$$\text{VSWR} = \frac{1 + \beta + \beta^2 + (1 + \beta) \sqrt{1 + \beta^2}}{\beta} \quad (6)$$

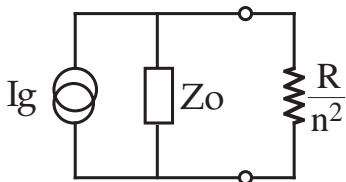


Figure 2: Cavity impedance at resonance as referred to the input.

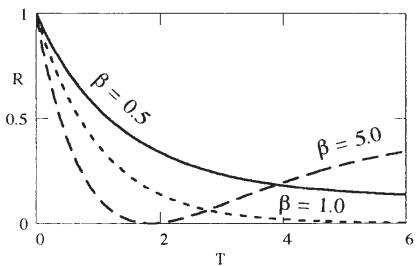


Figure 3:  $R$  vs  $T = t/2\tau$ .

**Very high  $Q$  cavities [3]** For cavities with  $Q_0 \gtrsim 10^5$ , the above approach becomes very difficult to apply. In that case one may make use of the long decay time involved for measurement of both  $\beta$  and  $Q$ . By observing the emitted power after the drive has been shut off (using a directional coupler) we have the relation at resonance,

$$\tau = Q_L/\omega \quad (7)$$

where  $\tau$  is the  $1/e$  time of the cavity energy. By observing the reflected power during filling one may also determine  $\beta$  from the ratio  $R(t)$  of reflected power to incident power,

$$R(t) = 1 - \frac{4\beta}{1+\beta} \left(1 - e^{-\frac{t}{2\tau}}\right) \left[ \frac{1-e^{-\frac{t}{2\tau}}}{1+\beta} + e^{-\frac{t}{2\tau}} \right] \quad (8)$$

The assessment of whether the cavity is matched, under or over coupled is easily made by viewing the time shape of  $R$  as shown in Fig.3 below for the three cases.

## References

- [1] H. Klein, CERN Acc. School RF Engineering for Particle Accelerators, CERN 92-03, vol.I, p.112
- [2] E.L. Ginzton, Microwave Measurements, McGraw-Hill (1957) Ch.10
- [3] H. Padamsee, T. Hays, J. Knobloch, Superconducting RF Cavities for Accelerators, Wiley (1998)

## 6.17 MAGNETIC MEASUREMENTS

### 6.17.1 Accelerator Magnets

A.K. Jain, P. Wanderer, BNL

**Basic formalism** Most magnets for particle accelerators use long coils with the current carrying conductors running parallel to the magnet axis. The field may be considered 2-D in the region far from the ends. The integral of the 3-D scalar potential over the entire length of a magnet (including the ends) also satisfies the 2-D Laplace's equation [1]. The field quality in accelerator magnets is expressed in terms of a harmonic description of the magnetic field in the aperture. The cylindrical components of the magnetic induction,  $B$ , may be expressed as [2]

$$B_r(r, \theta) = \sum_{n=1}^{\infty} \left( \frac{r}{R_{\text{ref}}} \right)^{n-1} (B_n \sin n\theta + A_n \cos n\theta) \quad (1)$$

$$B_\theta(r, \theta) = \sum_{n=1}^{\infty} \left( \frac{r}{R_{\text{ref}}} \right)^{n-1} (B_n \cos n\theta - A_n \sin n\theta) \quad (2)$$

where  $B_n$  and  $A_n$  are the  $2n$ -pole Normal and Skew components respectively.  $R_{\text{ref}}$  is an arbitrary reference radius, which should preferably be chosen to be roughly equal to the region of interest for the particle beam. The normal and skew components of the  $2n$ -pole field are sometimes denoted by  $B_{n-1}$  and  $A_{n-1}$ , often in the US literature. Also, the skew component is sometimes defined with a sign opposite to that used here [3].)

The Cartesian components of a 2-D field are most conveniently described in terms of a complex field [4],

$$\begin{aligned} \mathcal{B}(z) &= B_y(x, y) + iB_x(x, y) \\ &= \sum_{n=1}^{\infty} (B_n + iA_n) (z/R_{\text{ref}})^{n-1} \end{aligned} \quad (3)$$

where  $z = x + iy = re^{i\theta}$  is the point of interest in the complex plane.

**The “multipoles”** In practice one is interested in the magnitude of the various multipolar terms in relation to a reference field  $B_{\text{ref}}$ . Thus, we define the normal,  $b_n$ , and skew,  $a_n$  multipoles as

$$b_n = 10^4 \times \frac{B_n}{B_{\text{ref}}}, a_n = 10^4 \times \frac{A_n}{B_{\text{ref}}} \quad (4)$$

With the multiplying factor of  $10^4$ , the values are said to be in “Units”. The numerical values of the multipoles depend on the choice of reference radius and scale as  $R_{\text{ref}}^{n-1}$ . The reference

field  $B_{\text{ref}}$  should be chosen to match the convention used by particle tracking programs. It may be chosen as the strength of the main dipole field in a circular accelerator, but is generally chosen as the strength of the most dominant multipole term in the harmonic expansion. For example, one may use  $B_{\text{ref}} = |B_m + iA_m|$  for a  $2m$ -pole magnet. The normalized multipoles  $b_n$  and  $a_n$  scale as  $R_{\text{ref}}^{n-m}$  in this case.

**Transformation of coordinate axes** The coefficients  $B_n$  and  $A_n$  in Eqs.(1-2) are dependent on the choice of coordinate axes. If one needs to write an expansion of the field components in a coordinate frame which is rotated or displaced relative to the frame in which the coefficients are known, then one must obtain new coefficients [2] such that the magnetic field is correctly described by Eqs.(1-2) in the coordinate system of interest.

Rotation of axes Let  $B_n$  and  $A_n$  be the coefficients in the frame  $X-Y$ . The corresponding quantities in the new frame  $X'-Y'$  which is rotated by an angle  $\phi$  are, for all  $n \geq 1$ ,

$$\begin{aligned} B'_n &= B_n \cos(n\phi) - A_n \sin(n\phi) \\ A'_n &= B_n \sin(n\phi) + A_n \cos(n\phi) \end{aligned} \quad (5)$$

A rotation of axes, equivalent to an opposite rotation of the magnet, intermixes the normal and the skew components. Eq.(5) can be used to correct measurement data for any angular misalignment between the measurement frame and the magnet frame, or to estimate the effect of a roll of the magnet about the beam axis.

Displacement of axes If the axes in the new frame  $X'-Y'$  are parallel to the axes in the old frame  $X-Y$ , but the new origin  $O'$  has Cartesian coordinates  $(x_0, y_0)$  in the  $X-Y$  frame then the coefficients in the new frame are,  $n \geq 1$ ,

$$\begin{aligned} B'_n + iA'_n &= \sum_{m=n}^{\infty} (B_m + iA_m) \\ &\times \frac{(m-1)!}{(n-1)!(m-n)!} \left( \frac{x_0 + iy_0}{R_{\text{ref}}} \right)^{m-n} \end{aligned} \quad (6)$$

As seen from Eqs.(6), the  $2n$ -pole field coefficients in the displaced coordinate system are equal to the  $2n$ -pole field coefficients in the original coordinate system plus contributions from all the terms higher than the  $2n$ -pole. This effect is referred to as the feed down of harmonics. Eq.(6) is useful in estimating the effect of a misalignment of the magnet axis with the beam axis. They are also used to calculate the radial position of a

harmonic measuring system relative to the magnetic center of a magnet [5]. For example, to obtain the magnetic center of a quadrupole magnet, one would calculate  $(x_0, y_0)$  that would make the dipole terms ( $B_1$  and  $A_1$ ) zero. A first order calculation is generally adequate to calculate feed down from the most dominant term and for small offsets ( $< 0.1 R_{\text{ref}}$ ). Higher order calculations are often necessary for feed down from higher harmonics.

Inverted axes Sometimes a magnet is installed such that the positive  $X$ -axis defined for the accelerator, say  $X'$ , is opposite to the positive  $X$ -axis defined for the magnetic measurements. The origins and the  $Y$ -axis of the two systems are assumed to coincide. The transformation of the field expansion coefficients is

$$B'_n = (-1)^{n+1} B_n; \quad A'_n = (-1)^n A_n \quad (7)$$

Eq.(7) can also be used in some situations to estimate systematic errors in the measurement of harmonics by carrying out measurements from both lead and non-lead ends of the magnet. For example, for a dipole magnet the skew dipole term should exactly change sign in the absence of systematic errors. When reporting magnetic measurement results, the definition of the coordinate system must also be specified so that the appropriate harmonic transformations may be applied, as needed, by the users of the data.

**Magnet symmetries and allowed terms** In a  $2m$ -pole magnet, a rotation by  $\pi/m$  radians effectively reverses the current, or interchanges the north and the south poles. Thus, all field components reverse under this rotation. This implies that the only permissible terms in the expansion of the field from a  $2m$ -pole magnet are the  $[(2m)(2k+1)]$ -pole terms, where  $k$  is an integer including zero.

In addition to the fundamental rotation symmetry described above, the current distribution in a magnet may have additional symmetries resulting from an appropriate choice of coordinate axes. For example, if the  $X$ -axis lies in the midplane of a dipole magnet, the current distribution has a left-right anti-symmetry. Various normal or skew coefficients can be shown to vanish under such symmetries. The harmonic coefficients that become zero under various types of symmetries in the current distribution are listed in Tab.1.

Similar results for symmetries in the pole faces can be found in [1]. In a real magnet, unallowed harmonics may also be present due to

Table 1: Allowed Harmonics Under Various Symmetries in Magnet Current [6].  $k$  is any positive integer.  $\dagger$  Normal;  $\ddagger$  skew;  $L - R$ : Left-right symmetry;  $\bar{L} - \bar{R}$ : Left-right anti-symmetry;  $T - B$ : Top-bottom symmetry;  $\bar{T} - \bar{B}$ : Top-bottom anti-symmetry.

| Type                | Example  | Normal         | Skew           |
|---------------------|--|----------------|----------------|
| $L - R$             | Quadrupole $^\dagger$<br>Dipole $^{\ddagger\dagger}$ | $B_{2k-1} = 0$ | $A_{2k} = 0$   |
| $\bar{L} - \bar{R}$ | Dipole $^\dagger$                                    | $B_{2k} = 0$   | $A_{2k-1} = 0$ |
| $T - B$             | Dipole $^\dagger$                                    | All allowed    | $A_k = 0$      |
| $\bar{T} - \bar{B}$ | Dipole $^{\ddagger\dagger}$                          | $B_k = 0$      | All allowed    |

magnet leads, coil ends and construction tolerances. The specific unallowed harmonics present often give useful clues as to which symmetry is broken in the magnet.

**Rotating coils for measuring field quality** The field coefficients, or  $B_n$  and  $A_n$  are generally determined using a system of rotating coils. A coil essentially is a long loop of wire running parallel to the length of the magnet. The loop is rigidly supported by an insulating “coil form”. As this loop is rotated about an axis parallel (and as closely aligned as possible) to the axis of the magnet, a voltage signal is induced in accordance with the angular dependence of the field. The recording of this signal, generally at 128 or 256 equi-spaced angular positions, is triggered by a precision angle encoder. Fourier analysis of the induced signals leads to the determination of the field expansion coefficients. In addition, a rotating coil may also be equipped with a gravity sensor to obtain the field angle of the dominant term relative to gravity with a typical precision of 0.1–0.2 mrad. The voltage signals may be brought out of the rotating coil directly, in which case the rotation is limited to only a few turns in one direction. Alternatively, one could use slip rings to bring out the signals which allows the coil to rotate continuously. Depending upon the geometry of the loop, two main types of rotating coils are in common use (i) a tangential, or  $B_r$  coil and (ii) a radial, or  $B_\theta$  coil. The wires in the loop must be placed very precisely as per the design (typically within  $\sim 25\mu\text{m}$ ), and the as-built geometry must be calibrated for precise measurements.

**Tangential coil** The pick-up loop in a tangential coil lies in a plane normal to the radial vector through the center of the winding. See Fig.1. The

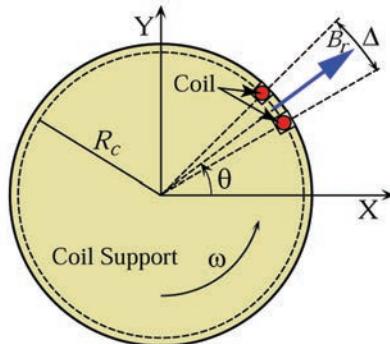


Figure 1: Cross-section of tangential coil.

flux through such a coil is determined by the radial component  $B_r$  of the field. For any angular position  $\theta$ , the total flux  $\Phi(\theta)$  through the winding can be obtained by integrating Eq.(1),

$$\Phi(\theta) = \sum_{n=1}^{\infty} \frac{P_t}{n} [B_n \sin(n\theta) + A_n \cos(n\theta)] \quad (8)$$

$$\text{where } P_t = 2NLR_{\text{ref}} \left( \frac{R_c}{R_{\text{ref}}} \right)^n \sin \left( \frac{n\Delta}{2} \right) \quad (9)$$

$N$  is the number of turns,  $L$  the length,  $R_c$  the radius and  $\Delta$  the opening angle of the coil. If the coil is rotated with an angular velocity  $\omega$  the voltage signal induced in the coil is

$$V(t) = \sum_{n=1}^{\infty} \omega P_t(n) [A_n \sin(n\psi) - B_n \cos(n\psi)] \quad (10)$$

where  $\psi = \omega t + \delta$ , and  $\delta$  is the angular position of the loop at  $t = 0$ . A Fourier analysis of  $V(t)$  gives the normal and skew harmonics  $B_n$  and  $A_n$ . Alternatively, one may integrate the voltage (e.g. to get rid of noise due to rotational speed fluctuations) and get a signal proportional to flux and use Eq.(8) to obtain the harmonics. While integrating the voltage has some advantages, one must correct for errors due to any drift in the integrator. This could be done by averaging data taken during clockwise and counter-clockwise rotations, or by comparing the flux at the start and the end of the rotation. Even when using integrators, it is important to keep the rotational speed fluctuations to a minimum to ensure proper correction for the integrator drift. Typically, a tangential coil designed to measure all harmonics of interest has a  $\Delta$  of 12–15° and has 20–30 turns. In order to obtain good

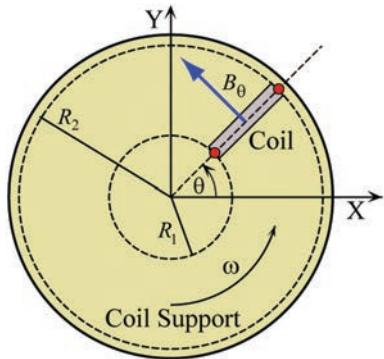


Figure 2: Cross-section of radial coil.

measurements of the higher harmonics,  $R_c$  should be at least equal to the radius of interest,  $R_{\text{ref}}$ , and preferably as large as the magnet aperture can accommodate.

**Radial coil** In a radial coil the plane of the pick-up loop is in the radial direction as shown by Fig.2. Typically, a coil is wound on a long, flat “bobbin” and then sandwiched between two halves of the coil form [7]. This technique may also be used to build tangential coils.

As seen from Fig.2, a radial coil is sensitive to the  $B_\theta$  component of the field. In this case,

$$\Phi(t) = \sum_{n=1}^{\infty} \frac{P_r(n)}{n} [B_n \cos(n\theta) - A_n \sin(n\theta)] \quad (11)$$

$$V(t) = \sum_{n=1}^{\infty} \omega P_r(n) [B_n \sin(n\psi) + A_n \cos(n\psi)] \quad (12)$$

where

$$P_r = NLR_{\text{ref}} \left[ \left( \frac{R_2}{R_{\text{ref}}} \right)^n - \left( \frac{R_1}{R_{\text{ref}}} \right)^n \right] \quad (13)$$

and  $\psi = \omega t + \delta$ ,  $\delta$  being the angular position of the loop at time  $t = 0$ . A Fourier analysis of  $V(t)$ , or the integral of  $V(t)$  gives the field coefficients [7, 8, 9, 10].

**Bucking and other issues** In practice, a rotating coil (either radial or tangential) also has several other windings with special geometries to “buck out” the most dominant harmonic, as well as the first order feed down terms in the signal. This is primarily done to eliminate spurious harmonics that may arise due to power supply ripple

and imperfections in the rotational motion of the coil. A detailed account of the theory of measurements can be found in [5], [8], [9], [10]. Bucking of the dominant harmonic also facilitates precise measurement of other harmonics which may be weaker by a factor of  $10^3 - 10^5$ . A design used for tangential coils at BNL has a total of five windings and includes a pair of dipole and a pair of quadrupole windings for bucking [11]. A bucking algorithm for using such a coil in magnets of different multipolarities is described in [5]. Practical radial coils also have additional windings designed to buck out the required terms for a given magnet type. The bucking can be achieved either by actually summing the signals from various windings and then measuring only the small resulting signal (*analog bucking*) or by recording the full signals from each of the windings and then carrying out the subtraction digitally (*digital bucking*). For digital bucking to work properly, the data acquisition system must have very high resolution (~ppm or better) and good linearity. A *bucking factor* (suppression factor for the most dominant term) of several thousand can be achieved with carefully adjusted analog bucking systems. Bucking factors of several hundred or more are desirable, although factors larger than several times  $10^4$  (e.g. with digital bucking) are not expected to provide additional improvement in the results for most accelerator magnets [10]. It is also important to precisely calibrate the radii and angular positions of the coil windings for accurate work. A procedure for calibrating geometric parameters of tangential coils having five windings can be found in [5].

With a good rotating coil system, it should be possible to measure harmonics with an accuracy of about  $\pm 0.1$  unit. Other factors that affect the measurement of harmonics include the sag of a long measuring coil due to its own weight, a tilt of the coil axis with respect to the magnet axis and various types of axial non-uniformity in the coil construction [5], [13].

**Measurement of magnetic center in multipoles** For installing quadrupoles, one needs to know the magnetic center and the angular roll of the magnet in reference to an external coordinate system. In addition, one needs the integrated gradient of the magnet. All these parameters of interest can be measured with the technique of the “stretched wire”. In this technique, a thin ( $\sim 100\mu\text{m}$  diameter) Cu-Be wire is passed through the bore of the

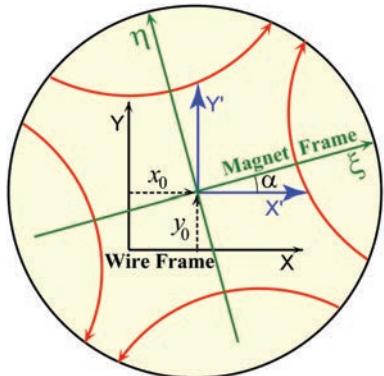


Figure 3: Coordinate system for the wire ( $X, Y$ ) and for the quadrupole ( $\xi - \eta$ ).

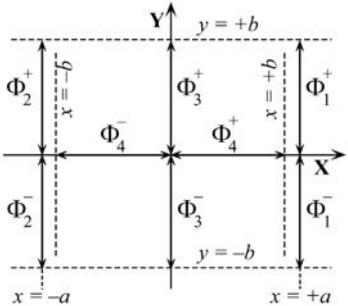


Figure 4: Movements of the wire during measurement.

quadrupole and held straight and parallel to the magnet axis. The wire can be moved in both horizontal and vertical directions using high precision translation stages. It forms an inductive loop together with an external return wire.

This loop is connected to an integrator to measure flux changes. Corrections must be applied for any drift in the integrator. A correction for the sag in the wire can be made by measuring as a function of tension in the wire.

Let  $X-Y$  be the external reference frame (defined by the motion of stages) in which we want to determine the magnetic center  $(x_0, y_0)$  of the quadrupole (Fig.3). The quadrupole's internal coordinate system is denoted by  $(\xi - \eta)$ , which is offset and rotated as shown in Fig.3. The wire is moved by a fixed distance,  $\pm b$ , along the vertical direction at three horizontal positions given by  $x = \pm a$  and  $x = 0$ , as shown in Fig.4. Similarly, the wire is moved in the horizontal direction

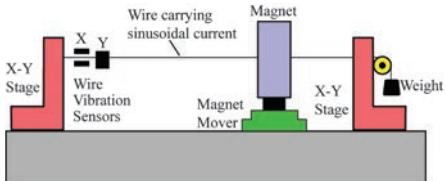


Figure 5: Flat coil for  $T$  measurement.

by the same distance  $b$  at  $y = 0$ . The corresponding fluxes measured by the integrated voltages are indicated in Fig.4.

The symmetric and anti-symmetric combinations of these fluxes are defined as

$$\Phi_j^S \equiv \Phi_j^+ + \Phi_j^-; \quad \Phi_j^A \equiv \Phi_j^+ - \Phi_j^- \quad (14)$$

For small roll angle  $\alpha$  and small offsets  $(x_0, y_0)$ , the integral of gradient  $G$ , the offsets, and the roll angle for a normal quadrupole are given by [7, 14]

$$b^2 \int G dl = \Phi_1^S = \Phi_2^S = \Phi_3^S = \Phi_4^S \quad (15)$$

$$x_0 \approx -\left(\frac{\Phi_4^A}{2\Phi_3^S}\right) b; \quad y_0 \approx -\left(\frac{\Phi_3^A}{2\Phi_3^S}\right) b \quad (16)$$

$$\alpha = \frac{1}{2} \arctan \left[ \frac{(\Phi_2^A - \Phi_1^A) b}{4\Phi_3^S} \right] \quad (17)$$

The magnetic centers of quadrupoles and correctors in RHIC have been determined by using a "survery antenna". This employs a non-rotating tangential coil with windings sensitive to specific harmonics. Analysis of voltage signals induced by sinusoidal excitation of the magnet at a few Hz gives the magnetic center [15]. Another method, suitable for quadrupoles at sufficiently high gradients ( $> 4$  T/m), uses scattering of polarized light through a colloidal suspension of ferric oxide particles to produce a visual indication of the magnetic center [16].

Fig.5 shows the setup for the vibrating wire technique [17], in which an AC current is passed through a stationary wire at a frequency matching one of the resonant modes of the wire. Any transverse magnetic field causes vibrations in the wire, which are detected and analyzed to estimate the field. Vibration amplitudes as a function of wire position give the field profile, from which the magnetic center can be derived. Measurement accuracy of  $\sim 5 - 10 \mu\text{m}$  in quadrupoles and sextupoles can be achieved [18].

**Integral transfer function of dipoles** The integral transfer function of a dipole is defined as

$$T \equiv \int_{-\infty}^{\infty} B_l(s) ds/I \quad (18)$$

where  $B_l(s)$  is the strength of the dipole field at axial position  $s$ , and  $I$  is the current in the magnet. The integral transfer function of long dipole magnets can be measured by either making a series of measurements with a short probe at several axial positions, or by using a single, long coil to cover the entire length of the magnet. The dipoles for HERA were measured with a system consisting of a NMR probe and a pair of Hall probes mounted at right angles to each other [7]. The NMR probe was used to obtain the absolute field strength in the straight section of the magnet. In the ends, the field varies rapidly with axial position, making it difficult to do the NMR measurements. The two Hall probes are used to obtain the magnitude and direction of the field in the ends. The measured field distribution can be integrated to obtain the integral transfer function. The accuracy of the integral is limited primarily by the positioning of the probe assembly. At HERA, the integral transfer function in 9 m long dipoles has been determined with an accuracy of 0.02% [7].

Another approach is to employ a long, non-rotating flat coil that covers the entire length of the magnet. The current in the magnet is ramped between two suitable limits and the signal output of the coil is integrated to obtain the change in flux. For a single flat coil of width  $w$  and with  $N$  turns , the integral transfer function is

$$T = \int_{t_1}^{t_2} V(t) dt/Nw [I(t_2) - I(t_1)] \quad (19)$$

where  $V(t)$  is the signal from the coil and  $I(t)$  is the current at time  $t$ . It is assumed here that the plane of the coil is normal to the field direction. This can be ensured using well designed fixtures if one has access to the pole tips [19], such as in conventional iron dominated magnets. The absolute accuracy of this method is limited by the knowledge of the coil width,  $w$ , although the reproducibility can be very good. If the width is calibrated against NMR/Hall probe measurements in a reference magnet, it is possible to obtain a precision of 0.02% [19].

The flat coil described above is not suitable when the plane of the coil cannot be easily referenced to the iron in the magnet, e.g. in superconducting magnets. To overcome this problem,

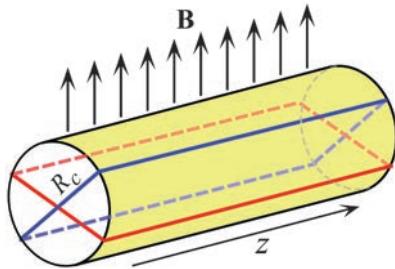


Figure 6: Pair of dipole windings for  $T$  measurement.

a non-rotating coil with two identical, orthogonal windings (Fig.6) has been used in RHIC [20].

If  $V_1(t)$  and  $V_2(t)$  are the signals from the two windings, the integral transfer function is

$$T = \frac{\sqrt{\left(\int_{t_1}^{t_2} V_1(t) dt\right)^2 + \left(\int_{t_1}^{t_2} V_2(t) dt\right)^2}}{2NR_c [I(t_2) - I(t_1)]} \quad (20)$$

where  $N$  is the number of turns and  $R_c$  is the radius of the windings. The expression for the integral transfer function is independent of the orientation of the field with respect to the coil. In practice, a slight angular dependence may be seen due to unequal radii and imperfect orthogonality of the two windings. One can calibrate the radii and the angle between the windings by recording the angular dependence in a magnet, and then use an appropriate formula for the integral transfer function. The absolute accuracy of this method is also limited by the knowledge of the radii of the windings, which could be calibrated against NMR/Hall measurements.

Eqs.(19, 20) assume that the field is a pure dipole field. The errors caused by higher harmonics in the field may not be negligible if one wants to achieve an accuracy of 0.02%. Corrections must be applied to account for the higher harmonics, or the coil geometry should be modified to eliminate sensitivity to the dominant harmonics, such as the sextupole [19]. Measurements with a rotating coil completely eliminate the effect of higher harmonics, and could be used to measure the integral transfer function in conjunction with a precise transporter mechanism to carry out an axial scan.

The stretched wire technique may also be used to measure the integrated field of any magnet. This method avoids the need to precisely calibrate a pick up coil as it relies on the precision of translation stages.

Table 2: Typical suppliers of instrumentation.

| Item                         | Manufacturer   |
|------------------------------|--|
| Motors                       | Rockwell Int, Reliance Electric, Cleveland, OH, USA<br>Nippon Seiko, Precision Prod. Div. Bloomingdale, IL, USA                    |
| Encoder                      | Dynamic Research Corp. Andover, MA, USA<br>BEI, Goleta, CA, USA  |
| Digital Voltmeter            | Agilent, Santa Clara, CA, USA  |
| Digital Integrator           | Metrolab, Geneva, Switzerland  |
| Current Transducer           | Hazemeyer, Holec System & Componenten, Amersfoort, Holland; Danfysik, Denmark  |
| Power Supplies               | Kepco, Inc. Flushing, NY, USA  |
| D/A Conv.                    | Iotech, Cleveland, OH, USA   |
| Slip Rings                   | Meridian Lab. Middleton, WI, USA   |
| Temp. Controller             | Omega Engr, Stamford, CT, USA  |
| Scanner                      | Cytec Corp. Penfield, NY, USA  |
| Gravity Sensors              | Spectron Glass & Electric, Hauppauge, NY, USA  |
| Hall Probes and Gauss-meters | Lake Shore Crytronics, Inc. Westerville, OH, USA<br>Group 3 Technology Ltd., Auckland, New Zealand<br>Sentron AG, Zug, Switzerland |
| NMR Probes and Gaussmeters   | Metrolab, Geneva, Switzerland  |

**Instrumentation** Tab.2 lists typical suppliers of the special instruments. The list is not comprehensive.

## References

- [1] W.C. Elmore, M.W. Garrett, RSI 25 (5) (1954) 480
- [2] A.K. Jain, CERN Accel. School on Magnetic Measurement and Alignment, (1997), CERN Report 98-05, p. 1.
- [3] P.J. Bryant, CERN Accel. School on Magnetic Measurement and Alignment (1992), CERN Report 92-05, p.52
- [4] R.A. Beth, J. Appl. Phys. 37 (1966) 2568
- [5] A.K. Jain, Harmonic coils, in Ref.[2], p. 175.
- [6] J.C. Herrera, Magnet Test Group Note No.110 (1981) BNL.
- [7] P. Schmueser, DESY HERA-P 92-1; CERN Acc. School (1992), CERN-92-05, p.240
- [8] W.G. Davis, NIM A311 (1992) 399
- [9] A. Devred, M. Traveria, Report CRYOMAG/94/08 (1994) Centre d'Etudes de Saclay
- [10] L. Walckiers, CERN Acc. School (1992), CERN-92-05, p.138
- [11] J. Herrera et al, PAC 89, p.1774
- [12] A.K. Jain, Proc. 9th Int. Magnet Measurement Workshop (1995), Vol.II
- [13] A.K. Jain, Tech. Note AD/SSC/Tech. No.99 (SSCL-N-784) (1992), BNL
- [14] J. DiMarco et al, IEEE Trans. Appl. Supercond. Vol.10(1), p.127 (2000)
- [15] A. Jain, Proc. 10th Int. Magnet Measurement Workshop (1997), Vol.II
- [16] J.K. Cobb, J.J. Murray, NIM 46 (1967) 99
- [17] A. Temnykh, NIM A 399 (1997) 185
- [18] A. Jain, in LER2010 Workshop, Jan. 12-15, 2010; <http://ler2010.web.cern.ch/LER2010/>
- [19] H.D. Glass et al, IEEE Trans. Magnetics, 32, No.4 (1995) p.3061
- [20] J. Herrera, unpublished work, BNL

### 6.17.2 Insertion Device Measurement

*S. Marks, R.D. Schlueter, LBNL*

**Requirements** Detailed magnetic measurements are essential to the task of commissioning an insertion device in a synchrotron radiation facility and preparing it for installation into a storage ring or free electron laser (FEL) facility . They are the basis for identifying and correcting magnetic field errors before installation and certifying magnetic performance. Magnetic field error requirements are driven by electron beam and radiation requirements; these,in turn, drive magnetic measurement requirements. Given a set of field error tolerances, measurement precision should be about an order of magnitude better in order to assure that requirements are met.

**Beam based requirements** Since the length of an insertion device is usually shorter than the local betatron wavelength, it is the full length field integral that is of most relevance. (Multipolar

expansions of these integral scan be found in [1] and Sec.6.17.1) As the beam lifetime is dependent on the dynamic aperture (Sec.2.3.9) of the lattice, which is in turn dependent on the field quality in all magnetic elements seen by the beam, it is important to measure the field in the insertion device throughout the volume defined by the specified dynamic aperture. The field integrals can then be inserted into a tracking code for assessment of the achieved dynamic aperture. The required constancy of the field integrals over the specified dynamic aperture usually are of the order of 100 G-cm [2, 3, 4]. Thus measurement precision of order 10 G-cm is in turn required.

**Radiation requirements** The radiation requirements for a wiggler are fairly modest. It is only important to insure that the peak field is what was expected, since the on-axis critical energy is related to the peak field by [5] (also Sec.3.1.3)

$$E_c [\text{keV}] = 0.665 E^2 [\text{GeV}] B_0 [\text{T}] \quad (1)$$

Since the wiggler radiation spectrum is broad, small peak-to-peak variations have negligible spectral effect; the effect on overall field integrals may be important, however. In contrast, spectral peaks from an undulator depend upon the coherence of radiation through the length of the device. Therefore, local field errors contribute to path length deviations corresponding to errors in the phase term  $e^{i\phi}$  in the radiation equation [6] (see also Sec.3.1),

$$\phi(t) = \omega \left[ t - \frac{\hat{n} \cdot \hat{r}(t)}{c} \right] \quad (2)$$

Phase errors correspond to path length errors,  $\Delta s(z)$ , which arise from the effect of magnetic field errors [7],

$$\begin{aligned} \Delta s(z) &= \int_{-\infty}^z \left\{ \left[ 1 + (x' + \Delta x')^2 + (y' + \Delta y')^2 \right]^{\frac{1}{2}} \right. \\ &\quad \left. - \left( 1 + x'^2 + y'^2 \right)^{\frac{1}{2}} \right\} d\xi = \Delta s_1 + \Delta s_2 + \Delta s_3 \end{aligned} \quad (3)$$

where

$$\Delta s_1 = x(z) \Delta x'(z) + y(z) \Delta y'(z)$$

$$\Delta s_2 = \int_{-\infty}^z [x(\xi) \Delta x''(\xi) + y(\xi) \Delta y''(\xi)] d\xi$$

$$\Delta s_3 = \frac{1}{2} \int_{-\infty}^z (\Delta x'^2 + \Delta y'^2) d\xi$$

$x, y$  and  $x', y'$  are the transverse positions and slopes of the particle trajectories,  $z$  the coordinate along the beam line and the  $\Delta$ 's represent changes from the ideal values due to the field errors.

The assessment of phase errors [8, 9, 10] requires local magnetic field measurements, either point measurements using point coils or Hall probes, or integral measurements over a length of a half period or less. Generally rms magnetic field errors  $\langle \Delta B_{\text{rms}} \rangle / \langle B_{\text{rms}} \rangle \approx 10^{-3}$  result in measurable degradation in spectral brightness. Therefore, point field measurement precision of the order  $10^{-4}$  or about 1 G or better for point measurements, is required to assure that spectral requirements are met. The ultimate characterization of the spectral performance is accomplished using a simulation code [11, 12, 13] that calculates spectral intensity using measured magnetic fields.

**Magnetic block measurements** Magnetic excitation of an insertion device can be from either current elements or permanent magnets (Sec.7.2.8), the latter has been the predominant choice..Permanent magnets provide higher fields than conventional conducting devices, except for very long periods. A subject of research in recent years has focused on the promise of superconducting devices [14]. Magnet block imperfections result in magnetic errors in an assembled device. The control of these errors requires establishing allowable variations in magnetic properties and the selective arrangement of blocks to minimize the effect of errors. Specifications for permanent magnet blocks generally include dimensional tolerances as well as allowable variations in the strength and orientation of the magnetic moment.

The three components of the magnetic moment vector are routinely measured with a Helmholtz coil. These measurements are used to screen blocks that do not meet specified requirements, and they provide the basis for sorting blocks for selective placement within the magnetic structure [15, 16]. Typical values for the primary magnetic moment component are of the order of 1.0 T. Allowed tolerances for variation in magnitude and orientation are about 1 %, so measurement precision should be about 0.1 %, or 10 G. For the purpose of selective placement the precision should be about 1 G.

Measurement of the magnetic moment with a Helmholtz coil provides an accurate and fairly easy way to characterize blocks. However, it has been recognized that the magnetic moment does not fully characterize a magnet block's effect on relevant magnetic errors [17]. Also of critical importance to the field quality are positioning

tolerances on the permanent magnet material, iron, and coils [3, 18] but as these do not entail magnetic measurements, they fall outside the present scope.

**Full length field integral measurements** We are generally interested in measuring  $\int_L Bdz$ , where  $L$  indicates integration over the full effective length of the insertion device along an axis coincident with or parallel to the beam axis. The magnetic field may be either  $B_x$  or  $B_y$ .

Full length field integrals are generally measured by use of an integral coil that extends through the entire magnetic structure and beyond each end to insure that it is past the fringe fields. Alternatively, the integral maybe derived from a full length scan of point measurements. The principle of integral coil measurements is based on

$$V = \dot{\Phi}N \quad (4)$$

where  $V$  is the induced voltage,  $\dot{\Phi}$  is the time rate of change of the enclosed magnetic flux, and  $N$  is the number of turns. To measure a static field, the coil is moved relative to the magnetic structure. If a coil is moved from a position or state where  $B = 0$  (or some other known field value) to the location aligned along the axis for which the integral is to be determined, the measured integral value is

$$\int_L \bar{B}_w dz = \frac{1}{Nw} \int V dt \quad (5)$$

where  $\bar{B}_w$  indicates average  $B$  over the coil width  $w$ . Note that as long as the coil is long enough to extend beyond the fringe field of the ends, the coil length does not enter the equation. Time integration of the voltage is generally accomplished with an analog or digital integrator [19]. Since an insertion device is generally several meters long, it is usually impractical to move a coil from a shield to the desired location within the magnetic structure in order to obtain an integral value. So alternatively, one can simply rotate the coil 180° about its axis and divide the voltage by 2. Often the gap is very small or the device is long, so it may be impractical to simply rotate a coil in-place. Nonetheless, since the integrated value is path independent, rotation can be accomplished by first sliding the coil out of the gap, then rotating and repositioning it within the gap. A precaution when using this approach: the integrated field excursion during the coil movement must not be large enough to saturate the integrator.

An insertion device represents a perturbation to a storage ring which may affect the tune and

beam-lifetime. Therefore, it is usually important to evaluate integrated multipoles and integral uniformity within the specified dynamic aperture. The on-axis field integral described previously corresponds to the integrated dipole. Higher order multipoles are derived from additional integrals measured off-axis but within the convergence radius. The multipoles may be determined from rotating coil measurements as is done for other accelerator magnets (Sec.6.17.1). Where it is impractical to use a rotating coil, the alternative is to measure the field integral at several lateral positions within the convergence radius. The multipole values can then be determined by a polynomial fit.

An integral coil can consist of a simple coil wound on a solid mandrel or wire loops that are held by fixtures and stretched through the magnetic gap[20]. Absolute integral measurements are then made by flipping the coil. Relative integral measurements are made by moving the coil laterally. Alternatively, a loop can consist of a section stretched through the insertion device gap with the remainder passing through a field-free or constant field region. A field integral is measured by translating the stretched portion laterally through the gap [21, 22]. The measured integral value corresponds to the flux lines passing through the area represented by the length of the stretched portion times the translation distance. Coil motion and integrator operation is often automated to accomplish faster data acquisition and better repeatability by insuring greater accuracy of motion.

With the use of low noise instrumentation [23], appropriately toleranced integral coils, and linear drift correction, integral measurement repeatability of the order of 1  $\mu\text{V}\cdot\text{s}$  can be achieved. The coil width and number of turns should be such that a signal corresponding to the required measurement precision is of this order or larger. For example, using a coil with 100 turns and a 1 cm width, a 1  $\mu\text{V}\cdot\text{s}$  signal corresponds to 1 G-cm. Green discusses a variety of issues concerning fabrication of integral coils [24].

Integration of point measurements The advantage of using long coils to measure full length field integrals, as described above, is that this is a direct measurement. However, an alternative is to integrate point measurement scans. The advantages are: (i) point measurements are required for evaluation of spectral quality anyway, so no

additional hardware is required; (ii) with point measurements the location of integral errors can be discerned for possible correction. In particular, some method for adjusting the integral contribution from the ends is generally incorporated in the magnetic structure. Point measurements allow the convenient separation of end integrals from that of the periodic section [25]; (iii) for FEL applications, a requirement is that the electron trajectory be within a specified tolerance over the full length. The trajectory is derived from the second field integral. In this application, the accuracy of the second integral function can be improved by using a full length coil field integral measurement to correct the full length field integral derived from point measurements.

The requirements on point measurement precision and variability may be more demanding than that required for evaluation of spectral properties. In addition, when using Hall probe measurements, small errors in the calibration can result in systematic integral offsets [26].

**Point measurement field scans** Point measurement field scans are performed on the nominal beam axis to determine spectral characteristics, and provide the basis for phase shimming [10, 27, 28]. Off-axis scans may also be performed to analyze integral information as described above. A scan density of 25 point measurements or more in a magnetic period allows a resolution of up to the 11th magnetic field harmonic and assures adequate determination of relevant field errors.

Hall probes are commercially available from a number of sources and are relatively inexpensive, convenient and reliable. However, several precautions must be observed:(i) They are temperature sensitive. This problem may be dealt with by calibrating temperature sensitivity and carefully monitoring temperature as measurements are taken. Another approach is to complete scans within a short period of time ( $\sim 1$  min.) and insure that ambient temperature does not drift significantly during a scan. This latter approach addresses relative field measurements within a single scan, which is far more important than absolute field measurements. Discrimination of field non-uniformities (i.e. relative values) to the order of 0.01% is achievable; absolute values need be resolved only to the order of 1%.(ii) Hall probes measure the field component perpendicular to the sensor; however, fields parallel to the sensor alter

the measurement via the planar Hall effect [29]. In many cases an insertion device has only one strong field component, usually  $B_y$ . For this case,  $B_y$  can be measured reliably as long as the sensor normal is aligned to within a few degrees relative to the  $y$ -axis. It may not be possible to reliably measure the  $B_x$  component for this case.

Hall probes may produce unreliable results when used for field scans of helical or elliptical devices where both  $B_x$  and  $B_y$  components are present; small coils or point coils present viable alternatives in such cases. Coils though, are not standard commodities and are generally custom made. There are two difficulties associated with the use of point coils. The first is the difficulty of providing enough turns to produce sufficient sensitivity while at the same time having a small probe area. One approach to this problem is to use flux balls instead of point coils [30]. A coil of this type is configured as windings on a sphere; this allows a larger surface for wire turns but still effectively produces a point measurement of the field at the sphere's center. An alternative approach is to recognize that point field values are not essential since even optical field errors occur as the result of the integration of trajectory deviations over a finite length. A coil which integrates fields over a quarter period is sufficient for assessing phase errors.

A second difficulty associated with coil measurements is the drift of integrators used to integrate coil voltage. Linear integrator drift may be easily dealt with by always measuring a known field (usually  $B = 0$ ) at the beginning and end of a scan. This measures linear drift and allows its subtraction from measurements. However, there is always some nonlinear component to integrator drift that is not accounted for; this results in measurement errors and non-reproducibility. The effects of integrator drift may be minimized by keeping scan times to the order of a minute or less. Both Hall probe temperature sensitivity and integrator drift motivate short scan times. These considerations must be balanced against noise rejection requirements. Generally the background electronic noise of an instrument increases non-linearly with increasing sampling rate. A duration of about one minute for a field scan of several meters generally represents a practical compromise between these requirements.

An alternative measurement technique has been developed and evaluated which uses a stretched conductive wire passing through an

insertion device gap[31, 32]. A current pulse is passed through the wire; deflection of the wire as a function of time is measured to determine magnetic field as a function of position along the wire.

## References

- [1] Reference manual for the Poisson/Superfish group of codes, LA-UR-87-126, LANL (1987) B13.3
- [2] G. Decker, APS Note, ANL (1992)
- [3] U10.0 undulator conceptual design report, LBNL Pub 5390 (1994)
- [4] R.D. Schlueter, Synchrotron Radiation Sources - A Technical Primer, World Sci. (1994) p.377
- [5] X-Ray data booklet, Center for X-ray Optics, LBNL Pub-490 Rev.(1986) 4
- [6] J.D. Jackson, Classical Electrodynamics, 3rd ed., Wiley (1999)
- [7] K. Halbach, in IDs, Wkshp. on 4th Generation Light Sources (1992)
- [8] B.L. Bobbs et al, NIM A296 (1990) 574
- [9] R.P. Walker, NIM A335 (1993) 328
- [10] B. Diviacco, R.P. Walker, NIM A368 (1996) 522
- [11] R.P. Walker, B. Diviacco, RSI 63 (1992) 1
- [12] Dejus, NIM A347 (1994) 61
- [13] C.X. Wang, D. Xian, NIM A288 (1990) 649
- [14] R. D. Schlueter et al., Synch. Radiat. News, vol. 17, no. 1, Jan/Feb. 2004
- [15] D.E. Humphries et al, RSI 66, 2 (1995) 1904
- [16] B. Diviacco, PAC 93, p.1590
- [17] J. Bahrdt et al, EPAC 96
- [18] R.D. Schlueter, K. Halbach, Symp. Mag. Tech. LBNL Pub 754 (1995) 55
- [19] M.I. Green, Magnetic Meas. Course, CERN Acc. School (1992)
- [20] D. Frachon et al, ANL/APS/TB-22 (1995)
- [21] R.P. Walker et al, PAC 95
- [22] D. Zangrando, R.P. Walker, NIM A376 (1996) 275
- [23] For example, High Precision Digital Integrator PDI 5025, MetroLab, Geneva Switzerland
- [24] M.I. Green, CERN Acc. School (1992)
- [25] S. Marks et al, SPIE Conf. (1993) 171
- [26] S. Marks, IEEE Trans. Magnetics, 30-4 (1994) 2435
- [27] S.C. Gottschalk et al, Synch. Rad. Instr. Conf. (1995)
- [28] J. Chavanne, P. Elleaume, Synchrotron Radiation News 8-1 (1995) 18
- [29] B. Berkes, CERN Acc. School (1992)
- [30] W.F. Brown, J.H. Sweer, RSI 16, 10 (1945) 276
- [31] R.W. Warren, NIM A272 (1988) 257
- [32] D.W. Preston, R.W. Warren, NIM A318 (1992) 794

## 6.18 HIGH POWER SWITCHES

### 6.18.1 DC Switches

*J. M. Sanders, T. Tang, M. A. Gundersen, G. Roth, U. Southern Calif.*

(See also Secs.6.11, 7.2.7.)

**Introduction** This section considers pulsed power switches for high voltage (1 kV - 100 kV), High current (1 kA - 100 kA) and fast switching time (ns -  $\mu$ s) applications. Because of the multiple simultaneous constraints ( $V, I, dI/dt$  and often repetition rate and life as well), this category of switches still includes gas phase as well as solid state switches.

**Gas phase switches** Gas phase switches have inherent advantages over solid state switches for certain forms of high pulsed power. Due to the robust nature of the gas/plasma, they are relatively immune to catastrophic failure due to circuit malfunctions. In addition, the use of a plasma to transport charge in these devices allows them to conduct currents orders of magnitude higher than most semiconductor devices - 100 kA current is not uncommon. Considered here are high pressure spark gaps (vacuum gaps, and vacuum tubes are not considered for these applications), and low pressure glow discharge pulsed power switches including the thyratron and pseudospark.

**High pressure spark gaps** A high pressure hydrogen spark gap is shown in Fig.1 [1]. The basic design incorporates a pair of trigger electrodes located between the main anode/cathode gap. A high voltage pulse is applied to the trigger electrodes, resulting in the injection of plasma into the anode-cathode gap. A spark breakdown then

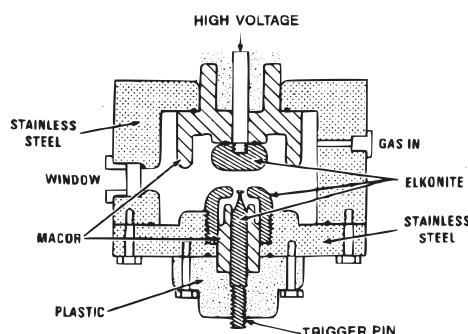


Figure 1: High power H<sub>2</sub> spark gap [1].

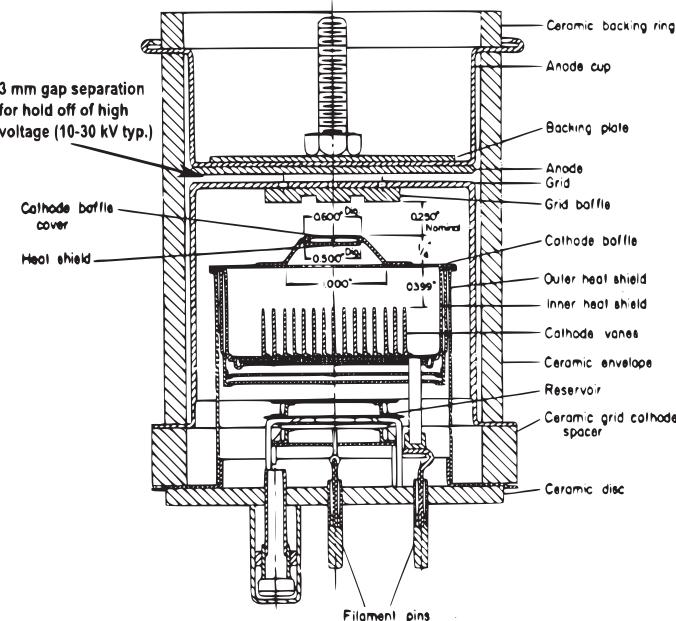


Figure 2: Schematic of a planar thyratron [7].

ensues between the anode and cathode. Spark gaps are filled with high pressure ( $> 1 \text{ atm}$ )  $\text{N}_2$ , air,  $\text{SF}_6$ ,  $\text{H}_2$  or even liquid media. Specialized spark gaps switch very high voltages ( $> 1 \text{ MV}$ ) and current (total charge  $> 100 \text{ coulombs}$ ).

A spark gap is typically triggered by either overvoltage the main gap or by internal auxiliary electrodes. Light can be used to assist the breakdown process in a spark gap and is one of the earliest reported experiments in modern physics [2]. Optical triggering using lasers has been reported, and such techniques can be used to reduce formative lag time and jitter [3, 4, 5]. The latter effect is quite important, as the typical spark gap  $\sim 25 \text{ ns}$  of jitter, unless special techniques are used. Jitter can be substantially reduced if very high overvoltage is used on the main gap in combination with UV preillumination ( $< 100 \text{ ps}$  has been reported.)

Spark gaps find application where high repetition rate is not important and where extremely fast switching times are required. Because the pressure inside a spark gap is fixed, the voltage range over which the device can operate is limited to  $\sim 1/3 V_{\max} - V_{\max}$ . In addition, due to the nature of the spark discharge, electrode erosion is significant for currents higher than a few kA, and

the lifetime of the tubes is relatively short if high current is conducted. If current is below a few hundred amps, spark gap erosion can be reduced significantly, and long life can be achieved.

**Thyatron** The idea of controlling a “power arc” with a grid electrode, named a “thyatron” from the Greek word for “gate”, was conceived and named by Langmuir [6]. Fig.2 shows a cutaway image of the typical ceramic hydrogen thyatron [7]. The thyatron is a closing switch, with closure achieved by grid control. A typical thyatron consists of a heated cathode, one or more triggering grids, an anode, and for the long lived high power switch, a heated reservoir of hydrogen. Its operation takes advantage of the interesting fact that the breakdown voltage of a system of plane parallel electrodes is not a linear function of the distance between the electrodes. Instead, when the breakdown voltage is plotted as a function of the product of the gas pressure and electrode spacing ( $p * d$ ) there is a minimum for a given  $p * d$ . Fig.3 illustrates this, usually referred to as a Paschen curve. In a typical high voltage hydrogen thyatron, the internal pressure of the thyatron and the spacing between the grid and anode is on the left-hand side of Fig.3.

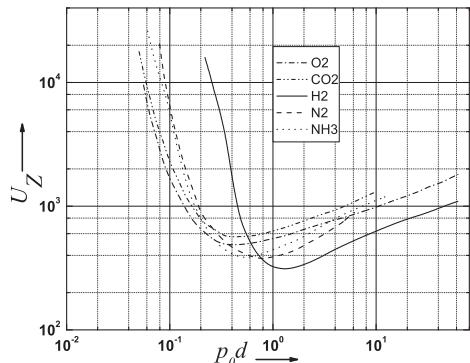


Figure 3: Typical Paschen curve [8].

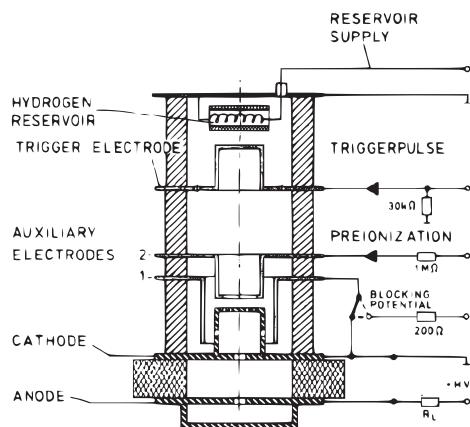


Figure 4: Structure of a pseudospark switch [9].

The thyratron typically recovers faster than a spark gap, allowing high repetition rates under some circumstances in excess of kHz. The lifetime of the switch depends strongly on the application but is generally good and can be  $10^9$  pulses. The jitter can be kept very low ( $< 1$  ns) with proper triggering.

A thyratron uses a heated cathode and also uses a heated reservoir of hydrogen to control the internal gas pressure. The heaters can consume several hundred watts of continuous power in high power tubes. The thyratron also has an inherent forward drop of a few hundred volts, which limits its efficiency at low working voltages. Most thytratrons have a minimum closure time of 20-50 ns, though closure times of 5 ns have been reported

with specially designed tubes. A thyratron is also sensitive to voltage reversal. Typical voltage reversals must be applied for recovery, yet must be kept to  $\sim 10\%$  of applied voltage, or else the lifetime of the tube can be considerably reduced.

**Pseudospark** A new group of high power low pressure thyratron-type switches that operate at thyratron-type pressures ( $\sim 0.1\text{--}0.5$  torr typical) includes the pseudo-spark (Fig.4 [9]) and BLT (Back-Lighted Thyratron) (Fig.5 [13]) [14]. The pseudo-spark was first proposed and studied at the University of Erlangen [15] for use as an ion and electron beam source. Initial experiments using a 40 gap device filled with 100 Pa hydrogen and an accelerating voltage of 100 kV produced a 70 keV electron beam with a current density of  $10^6$  A/cm<sup>2</sup>. In collaboration with a group from CERN the pseudo-spark was further developed as a fast triggered switch capable of  $> 100$  kA peak current [16]. The multi-channel pseudo-spark [17], a switch of pseudo-spark geometry with several discharge channels instead of one central hole, has been studied at the German French Research Institute, Saint Louis. Proton-antiproton collider experiments at CERN require focusing of very high current antiproton beams with a linear charged particle lens. A 400 kA pulse generator utilizing four pseudo-spark switches has been developed for this plasma lens. Over 100 kA of current was carried by each of the pseudo-spark switches used [18]. A “back of the cathode light triggered” thyratron (BLT) can be triggered by several optical methods and small single gap devices have achieved peak currents of  $\geq 50$  kA. The BLT employs the same cathode emission process as the pseudospark. This process is distinct from traditional thytratrons, involving a hollow cathode phase, followed by a superemissive phase [19]. High performance has been demonstrated in peak current ( $> 100$  kA), current rate of rise ( $> 10^{12}$  A/sec), switching precision, trigger efficiency, current reversal (100%), and recovery time. The BLT and pseudospark conduct, under long-life conditions, with a uniform glow plasma, as does a traditional thyratron. For reviews see [20, 21].

**Solid state switches** Though not capable of switching the same power densities as gas discharge switches, solid state switches are used increasingly in compact pulsed power systems that feature longer lifetimes and higher repetition rates than systems featuring gas discharge switches.

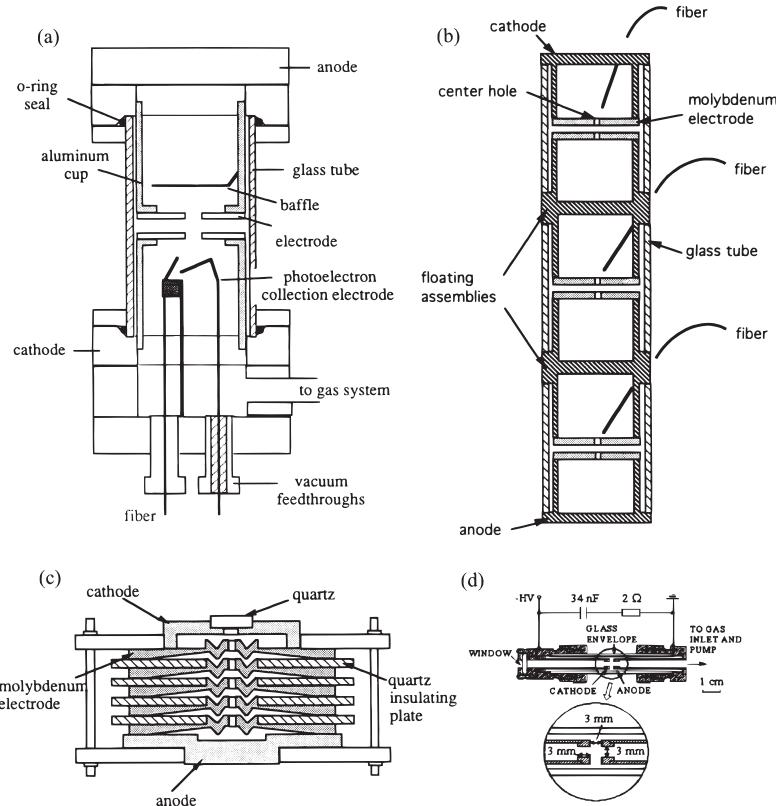


Figure 5: Structure of BLTs : (a) Fiber optic triggered BLT [10]; (b) Fiber triggered 3-gap BLT; (c) multi-gap BLT [11][12]; (d) Compact prototype mini-BLT switch and circuit schematic [13].

Principal solid state Si switches include the thyristor, IGBT, and MOSFET. Recent improvements in the production of SiC wafers have reduced the defects that have prevented the efficient production of SiC devices. New classes of power switches that utilize SiC's large bandgap are being developed to achieve higher voltage hold-off, lower on-resistance, and faster turn-on times [24]-[26].

**Thyristors** Since its development at Bell Labs in 1956, the thyristor (Fig.6) has taken many forms, including the silicon controlled rectifier (SCR), the gate turn-off (GTO) thyristor, the static induction thyristor (SITH), and the triode AC switch (TRIAC). The thyristor's switching characteristics combined with its high voltage hold-off, low voltage drop, and high operating current densities make it suitable for many high power applications, including the generation of high energy pulses. The thyristor's four layer pn structure en-

ables it to switch higher power densities than the other Si based switches, but carrier diffusion time in Si limits the current rate of rise, which leads to slower turn-on times than those featured by IGBTs and MOSFETs.

Si thyristors are commercially available from a number of manufacturers and feature AK hold-off voltages as high as 6.5 kV and average on-currents as high as 4.5 kA. To prevent turn-on di/dt failure, many of these devices are limited to rates of rise on the order of  $100 \text{ A}/\mu\text{s}$  for repetitive operation. Recent research has produced an experimental device made from a 4H SiC wafer that features a hold-off voltage of 9 kV and was shown to pulse a  $17 \mu\text{s}$  half-sine wave with an amplitude of 13.8 kA [24].

Thyristors are susceptible to different failure mechanisms. The most common of which is probably di/dt failure, which occurs when a high

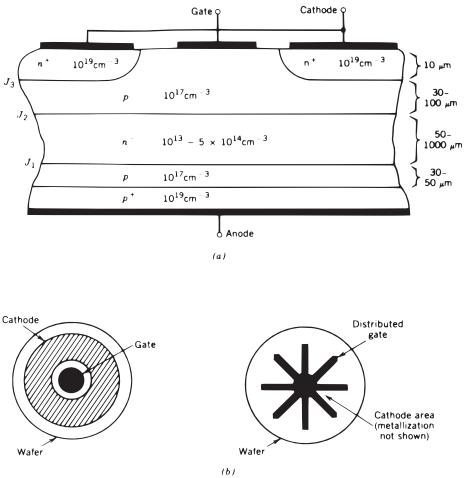


Figure 6: Cross section and gate/cathode layout of a high current thyristor [22].

current rise rate induces local heating in the initially narrow conduction region. Thyristors can also fail when transient reverse currents induce a voltage drop in the sub-cathode region that exceeds the reverse breakdown voltage of the gate cathode's diode junction. This failure mode is referred to as forced commutation. To avoid failure, snubber circuitry designed to reduce current commutation and slow turn-on times is typically employed. Other thyristor characteristics that may limit or adversely affect performance include latch-up, which prevents device turn-off by the gate signal, and switch-on dv/dt, which can trigger an undesired latch-up when quickly rising voltages are present at the anode or cathode.

**IGBT** The Insulated Gate Bipolar Transistor (IGBT) combines a MOSFET and a bipolar junction transistor (BJT) into one structure to create what is effectively a voltage controlled BJT. This device features desirable attributes for power electronics and pulsed power applications, such as low on-state voltage drop, high average current densities, and a voltage controlled gate, which simplifies gate driver circuitry. Commercially available IGBTs feature hold-off voltages as high as 4 kV and average currents as high as 400 A. Data from an experimental 4H SiC IGBT published in 2008 reported a hold-off voltage of 13 kV and operating current densities of 300A/cm<sup>2</sup> [25]. Fig.7 illustrates the structure of a typical IGBT, which contains an n-type MOSFET, a PNP

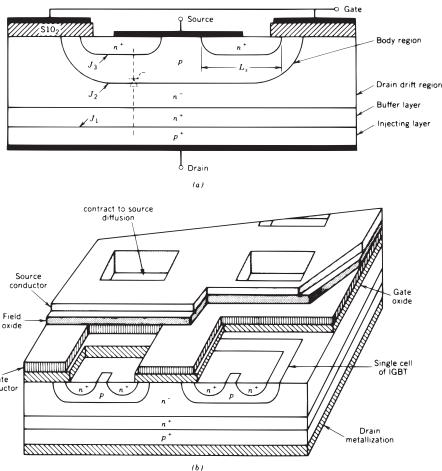


Figure 7: Cross section and perspective view of power IGBT [22].

BJT, an NPN BJT, and a JFET. The collector and emitter of the pn-n+p structure are the device's collector and emitter, or drain and source as they are referred to in Fig.7. The parasitic n+p-n structure combined with the pn-n+p structure makes a parasitic thyristor, which can potentially latch-up, if the device is not designed correctly.

Parallel IGBTs are commonly used in Marx bank configurations to achieve system voltage and current ratings that exceed those of an individual device. Variations on Marx bank configurations capitalize on IGBT's fast turn-on time and voltage controlled gate to generate arbitrary waveforms at high voltages and currents. The IGBT's voltage controlled gate, high repetition rate, and low cost make it a desirable solid state switch for many applications. The IGBT is outperformed by thyristors in applications requiring high power and by power MOSFETs in applications requiring fast turn-on.

**MOSFET** Power MOSFETs (Fig.8) are vertical MOS structures, usually n-channel, with VDS vs. IDS characteristics that are better suited for high current switching than those of standard, lateral MOSFET structures. Unlike thyristors and IGBTs, power MOSFETs are majority carrier devices and do not require minority carrier injection for operation. This translates into lower switching losses and faster turn-on, which makes high frequency operation of MOSFETs superior to other solid-state switches. With properly designed gate

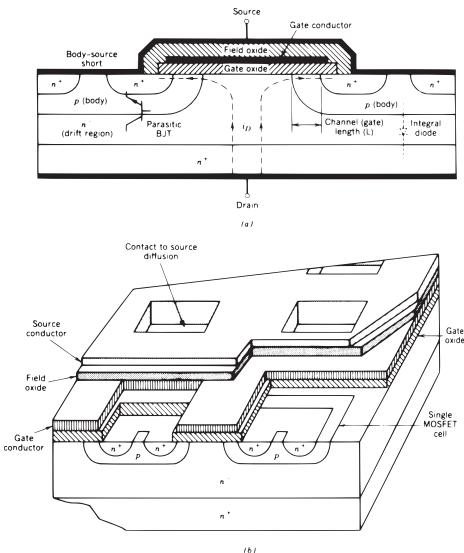


Figure 8: Cross section and perspective view of power MOSFET [22].

drive circuitry, power MOSFETs can be turned on in as little as 10 ns. The limitation on pulse length is the average power dissipated in the device's channel, which is determined by the load. If the impedance requires more current than can be provided by a single MOSFET, the devices can be parallelized. MOSFETs have a positive temperature coefficient, which indicates that their on-resistance increases with temperature. This makes parallelization easier because the parallel devices will self-regulate their current draw as the average dissipated power increases.

Commercial power MOSFETs are typically rated for drain-source hold-off between 50 V and 1.2 kV and average currents as high as 40 A. There are exceptions, including MOSFETs that are rated to hold-off 4 kV at no more than 1 A average current. The nominal on-resistance of Si power MOSFETs is typically about  $0.5 \Omega$ , which is higher than the on-resistance of an IGBT. SiC MOSFETs have recently been commercialized by CREE, Inc. These devices feature 1.2 kV hold-off, 20 A average current, and an on-resistance of  $18.7 \text{ m}\Omega$  [26]. The low on-resistance provided by SiC enables faster switching operation and faster risetimes. When properly cooled, these devices have been demonstrated to operate continuously at switching frequencies of 1 MHz.

**Diode Opening Switch** Diode opening switches, which are typically a type of drift step recovery diode (DSRD), are used to switch inductively stored energy on nanosecond and sub-nanosecond timescales. Their theory of operation was first described in 1985 by Igor Grekhov. In this and subsequent papers, Grekhov describes how a properly graded silicon pn junction can be pumped so that it transitions from a low impedance state to a high impedance state very quickly (i.e. on the order of 1 ns) [27]. Because the diodes must be pumped with a current to switch properly, they require auxiliary circuit components and switches to operate correctly. A number of architectures have been designed that use semiconductor power switches (MOSFETs, IGBTs) to switch energy into resonating circuits, MPC networks, or transmission lines to pump diode opening switches; the energy is subsequently switched into a load when the diode recovers [28]-[30].

Diode opening switches are used in systems capable of producing nanosecond and sub-nanosecond pulses with amplitudes between 1 kV and 100 kV. With proper cooling, these systems are capable of achieving repetition rates as high as 50 kHz.

**Magnetic Switch** A properly designed ferro- or ferrimagnetic core can be thought of as a switch that is triggered after a certain amount of magnetic flux has been applied. Magnetic materials feature a nonlinear relationship between magnetic flux and applied magnetic field, typically referred to as the hysteresis loop, that enables inductors and transformers to exhibit large changes in their effective inductance. Components that exhibit large changes in inductance can be utilized in circuits to switch and compress energy. Magnetic pulse compression (MPC) is a technique first introduced by Melville in 1951 as a means of compressing energy into a pulse [31]. A number of different architectures that employ magnetic switches and MPC have been developed since then [29], [30]. In a traditional MPC architecture, energy is resonantly transferred from one saturating stage to another. During the transfer, the period of the resonant network becomes shorter because the core's inductance drops by orders of magnitude upon saturation. Large factors ( $> 10^3$ ) of temporal compression are achieved as the energy propagates through the MPC network.

**Table 1 Approximate best specs. for different devices**

| Switch type          | Hold-off voltage (kV) | Peak current (kA) | di/dt (A/s)        | Risetime (ns) | Max Forward Drop (V)  | Max Ave. Power (kW) | Max Rep. Rate (kHz) | Trigger Control Method | Jitter (ns) | Turn-off cap.    | Life (shots)  |
|----------------------|-----------------------|-------------------|--------------------|---------------|-----------------------|---------------------|---------------------|------------------------|-------------|------------------|---------------|
| Spark Gap            | 100                   | 10 - 1000         | $10^{12}$          | 5             | 20                    | 20                  | 0.1 - 1             | HV Pulse               | <5          | No               | $10^6 - 10^8$ |
| Thyatron             | 125                   | 20                | $10^{12}$          | 20            | 150                   | 600                 | 40                  | HV Pulse               | 1           | No               | $10^{10}$     |
| Power Tube           | 200                   | 0.25              | $10^{11}$          | 100           | 20                    | 175                 | 50                  | HV Pulse               |             | Yes              | $10^{12}$     |
| Pseudospark          | 100                   | 5 - 100           | $10^{12}$          | 30            | 200                   | 600                 | 1                   | HV Pulse <sup>a</sup>  | 20          | No               | $10^9$        |
| Thyristor            | 1 - 9                 | 1 - 50            | $10^{10}$          | 1000          | 2                     | 10                  | 1                   | Current                | 5           | No               | $>10^{12}$    |
| MOSFET               | 1                     | 0.1               | $5 \times 10^9$    | 5             | 5                     | 0.5                 | $1 \times 10^3$     | Voltage                | 1           | Yes              | $>10^{12}$    |
| IGBT                 | 7                     | 1                 | $5 \times 10^9$    | 30            | 5                     | 1.25                | 20                  | Voltage                | 2           | Yes              | $>10^{12}$    |
| Saturable Inductor   | 130                   | 6                 | $10^{12}$          | 10            | (0.5 Ω) <sup>b</sup>  | Varies <sup>c</sup> | 20                  | Magnetic Flux          | 10          | Yes <sup>d</sup> | $>10^{12}$    |
| Diode Opening Switch | 1 - 10                | 1                 | $5 \times 10^{12}$ | 0.2           | (0.7-14) <sup>e</sup> | 0.5                 | 50                  | Current                | <1          | N/A              | $10^{12}$     |

a) Can also be triggered optically; b) Approximate parasitic resistance of windings. Voltage will also be dropped across the inductance associated with the windings on the saturated core; c) Depends on core size, geometry, and hysteresis loop; d) Here turn-off refers to using a bias current to reset the core; e) Here forward drop refers to the forward voltage of the diode, which accounts for power loss during the pumping phase.

Voltage hold-off, current handling capability, and switching speed are all functions of the core's constitutive parameters, geometry, and winding. Implementation of magnetic switches requires more design than implementation of semiconductor switches, but the added degrees of freedom empower the designer to tailor the switch to his/her application. Properly designed, impedance controlled MPC networks can switch pulses with amplitudes higher than 100 kV and durations less than 10 ns [32]. If the cores are properly cooled and kept below their Curie temperature, megawatt peak-powers can be generated at repetition rates above 20 kHz [33].

A relatively new amorphous, nanocrystalline material, commercially available from Metglas and MK Magnetics, features a unique combination of constitutive parameters that make it superior for MPC in some applications. Its high saturation flux density, high permeability, and lack of magnetostriction are all useful for efficiently achieving a high degree of temporal compression. Its losses at higher frequencies, however, limit its usefulness for achieving compression at low nanosecond or subnanosecond timescales. For timescales this short, low loss ferrites, such as NiZn, are more effective [34].

**Summary** Tab.1 shows approximate comparative parameters for the various switches. Many

switches, their applications to power modulators, and pulse generators are described in conference proceedings [23].

Some suppliers of high power switches, not an exhaustive list:

#### Spark gaps, thyatrons, pseudosparks

##### 1) MPD Components Inc.:

[www.mpdcomponents.com](http://www.mpdcomponents.com)

##### 2) E2V Technologies (Formerly EEV):

[e2vtechnologies.com](http://e2vtechnologies.com)

##### 3) Richardson Electronics: [www.rell.com](http://www.rell.com)

##### 4) L-3 (Formerly Litton Electron Devices):

[www.tw.l-3com.com](http://www.tw.l-3com.com)

##### 5) Pulsed Technologies Ltd., Russia:

[www.pulsetech.ru](http://www.pulsetech.ru)

#### Thyristors, IGBTs and MOSFETs

##### 1) IXYS (formerly DEI): [www.ixys.com](http://www.ixys.com)

##### 2) ABB Semiconductors AG:

[www.abb.com/semiconductors](http://www.abb.com/semiconductors)

##### 3) Powerex (a joint venture between GE & Mitsubishi): [www.pwrx.com](http://www.pwrx.com)

##### 4) International Rectifier: [www.irf.com](http://www.irf.com)

##### 5) Dynex Semiconductor: [www.dynexsemi.com](http://www.dynexsemi.com)

##### 6) Behlke Electronic (Semiconductor Switches):

[www.behlke.de](http://www.behlke.de)

##### 7) Intersil Corporation (former Harris Semiconductor): [www.intersil.com](http://www.intersil.com)

##### 8) Microsemi(formerly Advanced Power Technologies): [www.microsemi.com](http://www.microsemi.com)

## Sec.6.18: HIGH POWER SWITCHES

- 9) Mitsubishi Semiconductor:  
www.mitsubishichips.com
- 10) CREE: www.cree.com
- Magnetic Cores for MPC
- 1) Metglas, Inc.: www.metglas.com
- 2) MK Magnetics: www.mkmagnetics.com
- Diode Opening Switch Systems
- 1) Transient Plasma Systems:  
www.transientplasmasystems.com
- 2) FID GmbH: www.fidtechnology.com

## References

- [1] S.L. Moran, L.W. Hardesty, IEEE Trans. Electron Devices 38(4) (1991) 726
- [2] H. Hertz, Ann. d. Physik 31 (1887) 983
- [3] K. Harsch, H. Salzmann, H. Strohwald, Phys. Lett. A 55 (1975) 153
- [4] R.A. Dougal, P.F. Williams, J. Phys. D 17(5) (1984) 903
- [5] A.H. Guenther, J.R. Bettis, J. Phys. D 11 (1978) 1577
- [6] I. Langmuir, U.S. Pat. 1,289,823
- [7] S. Gouldberg, J. Rothstein, Advances in Electronics and Electron Physics, Academic (1961), Vol.XIV
- [8] M.J. Schonhuber, IEEE Trans. App. Syst. 88(2) (1969) 100
- [9] P. Bickel et al, IEEE Trans. Electron Devices 38(4) (1991) 712
- [10] C.G. Braun et al, IEEE Trans. Electron Devices 35(4) (1988) 559
- [11] T-Y Hsu et al, Proc. 7th IEEE Pulsed Power Conf. (1989) p.487
- [12] W. Bauer et al, Laser and Particle Beams 5(4) (1987) 581
- [13] C. Jiang, A. Kuthi, and M. A. Gundersen, Appl. Phys. Lett. 86(2), 24105 (2005).
- [14] G. F. Kirkman and M. A. Gundersen, Appl. Phys. Lett. 49, 494 (1986).
- [15] J. Christiansen, C. Schultheiss, Zeitschrift Fur Physik A-Hadrons And Nuclei 290(1), 35 (1979).
- [16] D. Bloess et al, NIM 205 (1-2) (1983) 173
- [17] G. Mechtersheimer, R. Kohler, J. Phys. E 20 (1987) 270
- [18] L. D. Menna et al, CERN-PS-84-13 11 (1984) 11
- [19] W. Hartmann, M.A. Gundersen, PRL 60 (23) (1988) 2371
- [20] K. Frank et al, IEEE Trans. Plasma Sci. 16(2) (1988) 317
- [21] M.A. Gundersen, NATO ASI Series B 219, ed. G. Schaefer, Plenum (1990)
- [22] N. Mohan, T.M. Undeland, W.P. Robbins, Power Electronics, Wiley (1995)
- [23] Proc. Int. Power Modulator Symposium, 1988~Present, available through IEEE-Customer Service Dept., Piscataway, NJ, tel. 1-800-678-IEEE
- [24] A. Agarwal, et al., Pulsed Power Conference 2009. PPC '09. IEEE (2009) 264.
- [25] M. K. Das, et al., Materials Science Forum 600 (2008) 1183.
- [26] C. Jonas, et al., Journal of Electronic Materials 37(5) (2008) 662.
- [27] I. V. Grekhov et al., Solid-State Electronics 28 (1985) 597.
- [28] V. M. Efanova, et al., Digest of Technical Papers, 11th IEEE International Pulsed Power Conference 2 (1997) 985.
- [29] T. Tang, et al., Dielectrics and Electrical Insulation, IEEE Transactions on 14 (2007) 878.
- [30] J. M. Sanders, et al., Dielectrics and Electrical Insulation, IEEE Transactions on (2011) In press
- [31] W.S. Melville, Proc. IEE 98 (1951) 185.
- [32] D. Birx, et al., Nuclear Science, IEEE Transactions on 30 (1983) 2763.
- [33] D.L. Birx, et al., UCID-18831, (1980).
- [34] J. E. Dolan, H. R. Bolton, and A. J. Shapland, Electronics Letters 33 (1997) 2049.

### 6.18.2 Ultra-High-Power Rf Switches *S. Tantawi, SLAC*

Fast Ultra-high-power microwave switches are required for active pulse compression systems. Research in this area concentrated on 4 types of switches: plasma switches [1], ferrite switches [2], bulk effect semiconductor switches [3], and recently electrically switched PIN diode active semiconductor windows [4]. Here, we will give the general design equations for any of these switches irrespective of the switching element, then we will show how these could be implemented with an optically controlled bulk effect semiconductor switch.

**Microwave control using a symmetric three port network** Consider the three port device shown in Fig.1. The device is composed of a basic *lossless* three port device with two similar ports, namely, port 1 and port 2. The third port is terminated so that all the scattered power from that port is completely reflected. However, the phase of the reflected signal from the third port can be changed actively. The material and the method by which its properties change so that its reflectivity would change are irrelevant to this analysis. This

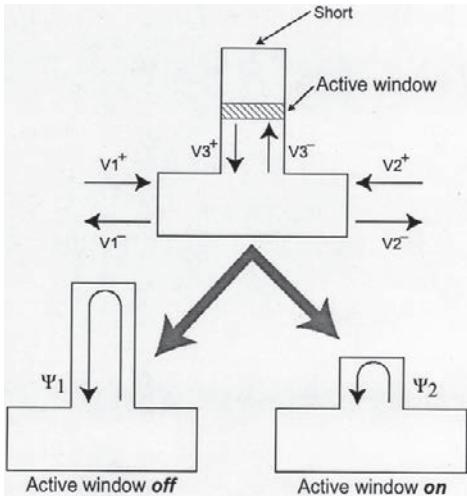


Figure 1: General structure of a high-power rf switch using 3-port network.

active window could be realized using any of the technologies mentioned above.

For any lossless and reciprocal 3-port network the scattering matrix is unitary and symmetric. By imposing these two conditions on the scattering matrix  $S$  of our device and at the same time taking into account the symmetry between port 1 and port 2, at some reference planes, one can write

$$S = \begin{bmatrix} \frac{e^{j\phi}-\cos\theta}{2} & \frac{-e^{j\phi}-\cos\theta}{2} & \frac{\sin\theta}{\sqrt{2}} \\ \frac{-e^{j\phi}-\cos\theta}{2} & \frac{e^{j\phi}-\cos\theta}{2} & \frac{\sin\theta}{\sqrt{2}} \\ \frac{\sin\theta}{\sqrt{2}} & \frac{\sin\theta}{\sqrt{2}} & \cos\theta \end{bmatrix} \quad (1)$$

Indeed, with the proper choice of the reference planes, this expression is quite general for any symmetric three-port network. The scattering matrix properties are determined completely with only two parameters:  $\theta$  and  $\phi$ . The scattered rf signals  $V^-$  are related to the incident rf signals  $V^+$  by

$$V^- = SV^+ \quad (2)$$

where  $V_i^\pm$  represents the incident/reflected rf signal from the  $i$ th port. We terminate the third port so that all the scattered power from that port is completely reflected; i.e.,

$$V_3^+ = V_3^- e^{j\psi} \quad (3)$$

The resultant, symmetric, two-port network, then, has the form  $S_{\text{two-port}} =$

$$\begin{bmatrix} \cos \frac{\zeta-\phi}{2} e^{j(\frac{\phi}{2}+\frac{\zeta}{2}+\alpha)} & j \sin \frac{\zeta-\phi}{2} e^{j(\frac{\phi}{2}+\frac{\zeta}{2}+\alpha)} \\ j \sin \frac{\zeta-\phi}{2} e^{j(\frac{\phi}{2}+\frac{\zeta}{2}+\alpha)} & \cos \frac{\zeta-\phi}{2} e^{j(\frac{\phi}{2}+\frac{\zeta}{2}+\alpha)} \end{bmatrix} \quad (4)$$

where the angle  $\zeta$  is given by

$$e^{j\zeta} = \frac{\cos\theta - e^{j\psi}}{\cos\theta e^{j\psi} - 1} \quad (5)$$

and  $\alpha$  is an arbitrary angle added to Eq.(4) so that the reference planes can be chosen at will. The signal level in the third arm is given by

$$|V_3^+|^2 = |V_3^-|^2 = \frac{\sin^2\theta}{3 - 4\cos\theta\cos\psi + \cos 2\theta} |V_1^+ + V_2^+|^2 \quad (6)$$

Eq.(6) shows that the signal level is independent of the angle  $\phi$ ; however, the angle  $\psi$  required to achieve given  $S$  parameters depend on the angle  $\phi$ . Solving (5) for the angle  $\psi$  and substituting into (6) one gets

$$|V_3^+|^2 = |V_3^-|^2 = \frac{1 + \cos^2\theta + 2\cos\theta\cos\zeta}{2\sin^2\theta} |V_1^+ + V_2^+|^2 \quad (7)$$

When the switch is *off*, i.e.  $|S_{12}| = 0$ , the angle  $\zeta = 2n\pi + \phi$  with  $n$  an integer. Assuming that the switch is looking at a matched load the third arm voltage level when the switch is *off* is given by

$$|V_3^+|^2 = |V_3^-|^2 = \frac{1 + \cos^2\theta + 2\cos\theta\cos\phi}{2\sin^2\theta} |V_1^+|^2 \quad (8)$$

Similarly when the switch is *on*, i.e.  $|S_{12}| = 1$ , the third arm signal level is given by

$$|V_3^+|^2 = |V_3^-|^2 = \frac{1 + \cos^2\theta - 2\cos\theta\cos\phi}{2\sin^2\theta} |V_1^+|^2 \quad (9)$$

Using (8) one can write an expression for the peak electric field at the surface of the active window when the switch is *off*,

$$E_{\max} = 2 \left( \frac{1 + \cos^2\theta + 2\cos\theta\cos\phi}{2\sin^2\theta} \right)^{1/2} \times \left| \sin \frac{\Delta\psi}{2} \right| \left( \frac{P_{\text{in}} Z_3}{A_3 G_3} \right)^{1/2} \quad (10)$$

where  $P_{\text{in}}$  is the constant level input power,  $Z_3$  is the wave impedance of the mode excited in the waveguide that forms the third arm,  $A_3$  is the cross sectional area of that guide, and  $G_3$  is a geometrical factors that depends on the mode and the

waveguide shape of the third arm, for TE<sub>10</sub> mode in rectangular guide  $G_3 = 1/4$ . Finally,

$$e^{j\Delta\psi} = \frac{(e^{j(\phi+\pi)} + \cos\theta)(1 + e^{j\phi}\cos\theta)}{(e^{j\phi} + \cos\theta)(1 + e^{j(\phi+\pi)}\cos\theta)} \quad (11)$$

Substituting from (11) into (10),

$$E_{\max} = \left( \frac{\sin^2\theta}{3 + \cos 2\theta - 2\cos(\theta - \phi) - 2\cos(\theta + \phi)} \right)^{\frac{1}{2}} \times 2 \left( \frac{P_{\text{in}}Z_3}{A_3G_3} \right)^{1/2} \quad (12)$$

When the switch is *on*, the losses on the surface of the active window is calculated with the help of (9) to be

$$P_l = \frac{1 + \cos^2\theta - 2\cos\theta\cos\phi}{2\sin^2\theta} \frac{R_s}{Z_3} P_{\text{in}} \quad (13)$$

where  $R_s$  is the surface resistance of the active element, given by  $(\omega\mu_0/2\sigma)^{1/2}$ . The conductivity  $\sigma$  depends on the level of excitation.

**An optical switch example** To actively change the angle of the reflection coefficient at the third port we place a piece of semiconductor material in the third arm. An external stimulus such as a laser light can induce an electron-hole plasma layer at the surface of the semiconductor, thus changing its dielectric constant. Therefore, the propagation constant of rf signals through the active arm changes; and consequently the coupling between the other two ports also changes.

Since silicon has a carrier life time that can extend from 1  $\mu$ s to 1 ms, it seems like a natural choice for this application. One can excite the plasma layer with a very short pulse from the external stimulus ( $\sim 5$  ns) and the device will stay in its new status long enough till all the rf signal is terminated. Indeed, this switch needs to have a very small amount of losses. Following classical arguments, one can show that the dielectric constant of a semiconductor material is

$$\epsilon = \epsilon_0\epsilon_r \left( 1 - \sum_i \frac{X_i}{1 - jZ_i} \right) \quad (14)$$

where

$$X_i = \frac{N_i e^2}{\epsilon_0\epsilon_r m_i^* \omega^2}, \quad Z_i = \frac{\nu_i}{\omega} \quad (15)$$

$\omega$  is the radial frequency of the rf signal,  $m_i^*$  is the effective mass of carrier  $i$  (electron, light hole and heavy hole),  $N_i$  is carrier density,  $e$  is the electron

charge, and  $\nu_i$  is the collision frequency. This latter quantity is related to the measured values of the dc mobility  $\mu_i$ ,

$$\frac{1}{\nu_i} = \frac{\mu_i m_i^*}{e} \quad (16)$$

Comparison between estimates of  $\nu_i$  for silicon to 10 GHz, an X-band switch for example, shows that  $Z_i \gg 1$ . Hence, one can show that the dielectric constant is given by the classical relation

$$\epsilon = \epsilon_0\epsilon_r \left( 1 - j \frac{\sigma}{\omega\epsilon_0\epsilon_r} \right) \quad (17)$$

where

$$\sigma = e \sum_i \mu_i N_i \quad (18)$$

which is the conductivity of the semiconductor.

To minimize the losses in the *off* state, i.e., when there is no plasma excited, we need to have a very pure semiconductor material such that the intrinsic carrier density is very small. In the *on* state, i.e., when the plasma layer is excited, the carrier density should be large enough so that the semiconductor acts like a good conductor and thus minimizing the losses.

At a carrier density  $10^{19}/\text{cm}^3$  silicon would have a conductivity of  $\sim 3.3 \times 10^3$  mho/cm. This is two orders of magnitude smaller than that of copper. However, it is high enough to make an effective reflector.

## References

- [1] A. Vikharev et al, Proc. 11th Advanced Accelerator Concepts workshop, 2004
- [2] S. Tantawi, PAC 2001 , p.1216
- [3] S.G. Tantawi, R.D. Ruth, A.E. Vlieks, M. Zolotorev, IEEE Trans. Microwave Theory & Techniques, Vol.45, No.8 (1997) 1486
- [4] F. Tamura, S.G. Tantawi, PRST-AB 5, no.6 (2002)

# Chapter 7. SUBSYSTEMS

## 7.1 PARTICLE SOURCES

### 7.1.1 Electron Sources

#### 7.1.1.1 Thermionic cathodes

*H.G. Kirk, BNL*

Thermionic emission occurs when electrons overcome the energy barrier (quantified as the work function  $\Phi_w$ ) and leave the surface of a material. This emission is facilitated by either raising the temperature of the emission material and/or by utilizing material with low values of the work function. This process was first quantified by Richardson[1]. Richardson's Law states:

$$J = AT^2 \exp\left[-\frac{e(\Phi_w - \Phi_s)}{k_B T}\right] \quad (1)$$

where  $J$  is the current density in  $A/cm^2$ ,  $A$  is a constant of proportionality  $\approx 60 A/(cm^2 \cdot K^2)$  for metals,  $T$  is the temperature of the emitter in  $^{\circ}K$ ,  $\Phi_w$  is the work function of the emitter material in  $eV$ ,  $\Phi_s$  is the reduction of the work function due to the Schottky effect[2], and  $k_B$  is the Boltzman constant. The work function  $\Phi_w$  is modified by the presence of an applied voltage to the emitting surface. This is known as the Schottky effect and is quantified by

$$\Phi_s = 0.012\sqrt{E_k} \quad (2)$$

where  $E_k$  is the applied external electric field in units of  $V/cm$ .

Typical work functions for commonly used emission materials are:

$$\Phi_w \approx \begin{cases} 4.5 \text{ eV, Tungsten} \\ 2.65 \text{ eV, Thorium-tungsten} \\ 1.2 \text{ eV, Oxides (BaO-SrO,etc)} \\ 2.5 \text{ eV, Hexaborides (LaB}_6, \text{CeB}_6) \end{cases} \quad (3)$$

The advantage of oxides such as BaO over other materials is that their low work functions allows for robust emission of electrons. Typical operating temperatures are in the range of 800 to

1000  $^{\circ}C$ . Oxides do, however, exhibit a substantial reduction in emissivity when exposed to poor vacuum conditions. This is referred to as cathode poisoning and as a result these types of cathodes are typically rendered useless after exposure to air until they are properly re-conditioned. The hexabromides, on the other hand, are more tolerant and can remain serviceable even after multiple interventions entailing, for example, cathode replacements. Typical operating temperatures for the hexabromides are in the range of 1200 to 1500  $^{\circ}C$ .

Dispenser cathodes have been developed to overcome the limitations experienced with the handling of low work function oxides. In this arrangement, the low work function material (barium for example) is impregnated within a mesh or sponge-like matrix of a host material such as tungsten. Upon heating, the embedded material evaporates and coats the outer surface of the cathode were its favorable work function allows for the robust emission of electrons. In an alternative configuration, the emissive material is stored in a reservoir situated just below the surface of the cathode. The maximum achievable current density for a non-relativistic beam is modified by the influence of the self-field of the particles leaving the cathode. This self-field repels the particles leaving the cathode surface thus inhibiting their exit. This is referred to as space-charge limited emission and has been quantified in Child's law[3] as:

$$I = PV^{3/2} \quad (4)$$

where  $I$  is the maximum achievable beam current in amps,  $V$  is the applied voltage between the cathode and gun anode, and  $P$  is the perveance of the gun which is dependent on the geometry of the acceleration chamber (gun). Note the clear advantage of establishing as high an electric field gradient as possible. This benefit is achieved by lowering the effective work function of the emitting material though the Schottky effects as well as

reducing space-charge limitations on maximum achievable beam currents. The emittance of the beam issuing from the thermionic cathode is determined by the size of the cross-section of the cathode emitting surface and the divergence of the beam leaving the surface. This divergence is dominated by the temperature of the cathode, hence:

$$\epsilon_{rms} = \frac{r_e}{2} \sqrt{\frac{k_b T}{m_e c^2}} \quad (5)$$

where  $\epsilon_{rms}$  is the rms emittance of the beam,  $r_e$  is the radius of the emitting surface,  $T$  is the temperature of the cathode, and  $m_e$  is the mass of the electron. Single crystals of Hexaborides have been successfully used to decrease the emittance of the emitted electron beam[4, 5]. Single crystals have the beneficial effect of reducing the divergence of the electrons issuing from the cathode surface as well as resulting in a more uniform distribution of the emitted electrons. Note that one is faced with a trade off in that a small radius is ideal for small emittances but larger radii are best for obtaining higher beam currents. Child's law will become an important consideration for this case.

## References

- [1] O.W. Richardson, *The Emission of electricity from hot bodies* (1921) Longmans, Green and Co., London
- [2] W. Schottky, Ann. der Phys. 57, 541 (1918) idem. 69 157 (1922)
- [3] C.D. Child, PR 32, 492 (1911)
- [4] K. Togawa, et al., *CeB<sub>6</sub> electron gun for low-emittance injector* PRST AB 10, 020703 (2007)
- [5] K. Kasamook, et al, APAC07, (2007) p183

### 7.1.1.2 Photocathodes (unpolarized)

*D.H. Dowell, SLAC*

The cathode types commonly used in RF guns are metal and semi-conductor photo-emitters. Metal cathodes are commonly used in high field guns where the high field both increases the QE via the Schottky effect and rapidly accelerates the beam. Since high rf fields are more easily achieved at high rf frequency, the temporally fast emission of metal cathodes make them a good match with s-band and higher frequency guns. In addition metals are more robust in the poor vacuum and the high voltage environment, and most guns using metal cathodes do not have load locks. The

high peak fields of these guns usually limit them to low duty factor operation. High duty factor guns require cathodes with high quantum efficiency at visible wavelengths. These requirements result from the average power limit of commercially available laser technology and the increased danger of optical damage to the cathode and the laser optics at UV wavelengths. Three important characteristics of photocathodes are quantum efficiency, thermal or intrinsic emittance and response time.

**Photocathodes and Quantum Efficiency** The laser average power in terms of the average beam current, the photon wavelength and the QE is given by

$$P_{laser}(\text{watts}) = 1240.4 \frac{I(\text{amperes})}{QE \times \lambda(\text{nm})}$$

QE's widely vary with material beginning with the lowest from metals (QE from  $10^{-5}$  to  $10^{-4}$ ), followed by enhanced metal or coated metal cathodes (QE from  $10^{-4}$  to  $10^{-3}$ ) and then the high sensitivity semi-conductor cathodes (QE from  $10^{-3}$  to  $10^{-1}$ ). Sophisticated hybrid cathodes are being developed which have QE's greater than 100% due to production of secondary electrons [1]. In order to achieve a realistic, commercially available laser system and to limit cathode heating and damage, the average power on the cathode should be kept below  $\sim 100$  watts. This means the QE needs increase with the higher average current. This average power limits metal cathodes to an average current of  $\sim 1$  mA. For currents between 1 to a few tens of mA, metal cathodes enhanced either by a coating such as CsBr [2] or increased emission due to a structured surface [3], are practical. As are semi-conductor cathodes with a protective coating [4]. The very high current regime (100 to 1000 mA) requires QE's of 0.01 to 0.10 or higher with sensitivity at visible wavelengths. Semi-conductor cathodes which have these high QE's can be classified as either Positive Electron Affinity (PEA) or Negative Electron Affinity (NEA) cathodes. The PEA alkali cathodes such as K<sub>2</sub>CsSb is actively being researched for use in high duty factor guns [5] and CsTe is successfully being used in the L-band gun for the FLASH FEL [6]. NEA cathodes can have much lower intrinsic emittance than PEA cathodes but this thermal component of the emission results in a slower temporal response as described in Ref.[7], whereas the response time of most PEA cathodes is less than one picosecond.

The properties of several photocathodes are given in Tables II and III of Ref [8]. There is also interesting recent work of photoemission using short ( $\sim 100$  fs) laser pulses to enhance multi-photon electron emission [13]. Multi-photon processes dominate when the laser pulse length is nearly the same as the electron emission time.

**The Thermal Emittance** In addition to having a good QE, the cathode also needs to have a low thermal or intrinsic emittance. Applying the three-step model of photoemission [9] gives the following general expression of the intrinsic emittance for photoemission [10]

$$\frac{\varepsilon_{n,metal}}{\sigma_x} = \sqrt{\frac{E_{excess}}{3mc^2}}$$

The intrinsic emittance has units of microns/(mm-rms) and a rather simple dependence upon the excess energy,  $E_{excess}$ , defined as the difference between the photon energy,  $\hbar\omega$ , and the energy required to extract an electron.  $\sigma_x$  is the rms transverse beam size and  $mc^2$  is the electron mass energy. The excess energy for a metal cathode is the difference between the photon energy, the material work function,  $\phi_W$ , and the Schottky energy,  $\phi_{Schottky}$ , [11]

$$\begin{aligned} E_{excess} &= \hbar\omega - \phi_W + \phi_{Schottky} \\ &= \hbar\omega - \phi_W + 0.0379\sqrt{\beta E(MV/m)}eV \end{aligned}$$

Here  $\beta$  is the field enhancement factor and  $E$  is the applied RF field. The intrinsic emittance for semi-conductors is more complicated due to additional internal processes such as electron-phonon scattering. However, similar to metal cathodes, the excess energy in terms of the band gap energy,  $E_G$ , and electron affinity,  $E_A$ , can be approximated by

$$E_{excess,semi} = \hbar\omega - E_G - E_A + \phi_{schottky}$$

These simple relations allow comparison of the intrinsic emittances for the various cathodes used in rf guns as is done in Ref.[8].

A notable exception to this simple view is the intrinsic emittance of GaAs and other NEA cathodes which exhibit both prompt and delayed emission with much lower thermal emittance for the delayed emission [7].

**The Lowest Possible Emittance** The theory for the intrinsic emittance indicates that the smaller the laser beam size on the cathode the lower the emittance. The smallest laser beam size which

can emit a charge,  $Q$ , is given by the space charge limit. The space charge limit for short pulse photoemission occurs when the surface charge density equals the applied rf field,  $E_{rf}$ , leads to an expression for the intrinsic emittance giving the lowest possible emittance from a photocathode,

$$\varepsilon_{lowest} = \sqrt{\frac{Q}{12\pi\varepsilon_0 mc^2} \frac{E_{excess}}{E_{rf}}}$$

For the case of a copper cathode having a 4.6 eV material work function illuminated with a 4.86 eV photon in an electric field of 50 MV/m at the time of emission, the lowest possible emittance for 1 nC is 0.27 microns. A similar relation can be derived for semi-conductors [12].

**Vacuum Issues** The vacuum requirements vary widely for the different cathode types but in general all photocathodes require an excellent vacuum. Even the metal cathodes, such as copper, can exhibit significant QE reductions when operated at high repetition rates. Generally metal cathodes can be used at total vacuum pressures of  $10^{-9}$  torr, while the positive electron affinity semi-conductors require  $10^{-10}$  torr and an NEA like GaAs requires low  $10^{-11}$  torr [8]. Trace amounts of an oxidizing contaminant, such as oxygen or water, can actually increase the QE by “activating” the material, but in larger doses will kill the cathode [5].

## References

- [1] J. Smedley, et al., Diamond Electronics—Fundamentals to Applications II, Mater. Res. Soc. Symp. Proc. 1039, Warrendale, PA, 2007, 1039-P09-02; Chang, X. et al PAC 09, (2009) p 693
- [2] Juan R. Maldonado et al., PRST AB 11, 060702 (2008); K. Nemeth et al., PRL 104, 046801 (2010)
- [3] J.G. Endriz, Applied Physics Letters, 25, 1974, p.261; F. Sabary et al., Applied Physics Letters, 1991. 58(12): p. 1230-1232; J. Le Perche et al., PRL, 2008. 100(6): p. 066408-4; S.E. Irvine, A. Dechant, A.Y. Elezzabi, PRL, 2004. 93(18) p.184801; P. Dombi, P. Rácz, Optics Express, 2008,16(5): pp. 2887-2893
- [4] A. Buzulutskov et al, NIM A400(1997)173
- [5] D. H. Dowell et al., NIM A356(1995)167-176; K.L. Jensen, et al., J. Appl. Phys., 104 (2008) 044907
- [6] M. Miltchev et al, FEL05 (2005) p.563
- [7] I.V. Bazarov et al, JAP, 103 (2008) 054901; PAC07, p. 1223

## Sec.7.1: PARTICLE SOURCES

- [8] D.H. Dowell et al, NIM A622(2010)685-697 references therein
- [9] W.E. Spicer, A. Herera-Gomez, SLAC-PUB-6306
- [10] D. H. Dowell, J.F. Schmerge, PRST AB 12, 74201(2009)
- [11] D. H. Dowell et al., PRST AB **9**, 63502(2006)
- [12] I.V. Bazarov, B.M. Dunham, C.K. Sinclair, PRL, **102** (2009) 104801
- [13] P. Musumeci et al., PRL **104**,084801(2010)

### 7.1.1.3 Photocathodes (polarized)

*T. Maruyama, SLAC  
M. Poelker, JLAB*

**Cathode materials used in photoemission sources** III-V semiconductor photoemission sources of polarized electrons, introduced in the mid-1970s [1], have been the most successful in high current applications such as injectors for electron accelerators. Direct band-gap III-V semiconductor materials include gallium arsenide and GaAs-related compounds. The materials are either in bulk form, or layered. Layered structures are grown by MBE or MOCVD processes. Strain is introduced by depositing an epitaxial layer on a substrate with a different lattice constant. Table 1 lists a representative set of these materials. Quantum efficiency (QE) is the yield of electrons per incident photon.

**Cathode preparation** Cathode materials must be treated for negative electron affinity (NEA) for efficient photoemission. The treatment consists of heat cleaning in an ultrahigh vacuum environment, followed by deposition of cesium and oxygen or fluorine (“activation” [7]). QE is monitored and maximized. QE is typically observed to decay slowly with time, and recoating with cesium (“rescesiation”) after loss of QE can restore efficient photoemission.

**High-gradient doping** A high *p*-type dopant concentration is required to achieve high QE and high charge limit. However, increasing the doping level leads to some spin depolarization. To increase the doping level without spin depolarization, a high-gradient-doping technique is employed: the active layer doping level is reduced to  $5 \times 10^{17} \text{ cm}^{-3}$  while increasing the concentration of a 10-nm surface layer to  $5 \times 10^{19} \text{ cm}^{-3}$ . High-gradient-doped GaAsP-GaAs strained-superlattice wafers can be commercially grown by SVT Associates [8].

**Design of a photoemission source** The entire source volume must be bakeable to ultrahigh vacuum (total pressure  $< 10^{-11}$  torr). The design must allow for cathode preparation in situ or nearby. Ability to insert and remove the cathode through a vacuum valve (“load lock”) is desirable. See Ref.[7]. Alternative designs permit the load lock section to remain at ground potential during source operations [9]. Extracted beam striking walls quickly degrades vacuum, so optics design must include beam stay clear requirements. The laser wavelength is set by the choice of cathode material. The laser power and pulse structure is set by the accelerator parameters. The laser beam should be circularly polarized and reversible.

**Current limits** Maximum current available is set by one of the following factors, whichever is lower.

(i) laser pulse energy and QE value; For a laser pulse of energy  $W$  [joules], the number of emitted electrons is given by

$$Ne = 5.03 \times 10^{13} \text{ QE}[\%]W[J]\lambda[\text{nm}]$$

(ii) space charge limit [10]; The space charge limit (Child’s law, Sec.7.1.1.1) is a current limit which scales as  $kV^{3/2}$ , where  $k$  is the permeance. The permeance depends on the geometry of the electrodes, and must be evaluated or measured for each gun geometry. The space charge limit of a gun also depends on the spatial and temporal characteristics of the extracted beam with higher current limits obtained for smaller beams [11] and short pulses compared to the transit time of the bunch within the cathode/anode gap [12].

(iii) cathode charge limit [13]; The charge which can be drawn in short pulses is observed to saturate as the incident laser pulse energy is increased. This limiting process is related to charging of surface states, which increases the work function, temporarily choking off the flow of current. This limitation is significantly reduced for dopant densities at the surface of  $\geq 10^{19} \text{ cm}^{-3}$  [14].

**Polarization reversal** Illuminating the cathode with circularly polarized light gives longitudinally polarized electrons; reversal of the circular polarization reverses the longitudinal electron

Table 1: Photocathode Materials.

| Type                    | Pol (%) | $\lambda$ (nm) | QE(%) at $\lambda$ | Ref. |
|-------------------------|---------|----------------|--------------------|------|
| bulk GaAs               | 40      | 750            | 10                 | [2]  |
| strained InGaAs         | 70      | 985            | .01                | [3]  |
| strained GaAs           | 85      | 850            | 0.2                | [4]  |
| superlattices           |         |                |                    |      |
| GaAsP-GaAs              | 90      | 780            | 1                  | [5]  |
| $Al_xIn_yGaAs-Al_zGaAs$ | 70-90   |                | up to 1            | [6]  |

polarization (helicity). Passing a laser beam through a linear polarizer followed by a quarter-wave plate (suitably oriented) establishes circularly polarized light. Rotation of the quarter-wave plate by 90 degrees reverses the sign of the polarization. Rapid reversals are desired in experiments measuring small effects, such as parity violation. Pockels cells are used for rapid reversals, and subtle effects in the Pockels cell must be controlled to avoid affecting the electron beam.

**Helicity-correlated asymmetry** Due to the built-in strain and anisotropic strain relaxation, QE anisotropy is present in photoemission when optically pumped with linearly polarized light [15]. The existence of such a QE anisotropy may lead to helicity dependent intensity variations if there exists a small, unintentional, linear component in the circularly polarized pump beam. Procedures to null these and other effects have been developed [16].

## References

- [1] D.T. Pierce et al, RSI 51 (1980) 478
- [2] R.L. Bell, Negative Electron Affinity Devices, Clarendon Press (1973) ; D.T. Pierce in F.B. Dunning and R.G. Hulet, eds., Atomic Molecular, and Optical Physics: Charged Particles, Academic Press (1995), p. 1
- [3] T. Maruyama et al, PRL 66 (1991) 2376
- [4] T. Nakahishi et al, PL A158 (1991) 345 ; H. Aoyagi et al, PL A167 (1993) 415 ; T. Maruyama et al, NIM A492 (2002) 199
- [5] T. Maruyama et al, APL 85 (2004) 2640 ; T. Nishitani et al, JAP 97 (2005) 094907
- [6] T. Nakanishi et al, AIP Conf. Proc. 421 (1998) p. 300 ; Yu. A. Mamaev APL 93 (2008) 081114
- [7] R. Alley et al, NIM A365 (1995) 1
- [8] SVT Associates, Eden Prairie, Minnesota 55344 <http://www.svta.com>
- [9] M. Breidenbach et al, NIM A350 (1994) 1; N.H. Papadakis et al, Polarized Beams and Gas Targets, World Scientific (1996) p. 323; K. Aulenbacher et al, NIM A391 (1997) 498; P. Adderley et al., PRST AB 13, 010101 (2010)
- [10] S. Humphries, Charged Particle Beams, Wiley (1990), Ch.5
- [11] Y.Y.Lau, PRL, 87, 278301 (2001)
- [12] L.K.Ang and P.Zhang, PRL 98, 164802 (2007)
- [13] M. Woods et al, JAP 73 (1993) 8531
- [14] K. Togawa et al, NIM A455 (2000) 118; T. Maruyama et al, NIM A492 (2002) 199
- [15] R.A. Mair et al, PL A212 (1996) 231
- [16] T.B. Humensky et. al., NIM A 521, 261 (2004)

### 7.1.1.4 DC thermionic guns and preinjectors

*H.G. Kirk, BNL*

Electron beams can be formed by producing a potential difference between the thermionic emitter (cathode) and a plate (anode) with an orifice to permit the electron beam to exit the chamber. Frequently the emission of the electrons from the cathode can be further tuned by utilizing a grid between the cathode and anode (a classic triode arrangement) or by placing the cathode in a Wehnelt cylinder assembly. A Wehnelt cylinder has its axis coinciding with the desired beam path and the end is terminated with a closed surface with an aperture to allow the beam to exit. The emission of the electrons from the thermionic emitter can be fine tuned by placing a shaping negative bias on the grid or Wehnelt cylinder. The potential difference between the cathode and anode typically spans the range of 50 to 500 kV . Although this is referred to as a DC gun,, the high voltage is typically pulsed so that the electron beam's temporal length is normally 1 to several  $\mu$ s. A typical arrangement for a DC gun is shown in Fig.1.

Following the beam's exit from the gun's anode plate, a velocity modulation can be imparted upon the beam by means of radio frequency (rf) cavities. Bunches are formed by imparting an energy loss at the head of the electron stream followed by an energy gain at the tail of the stream.

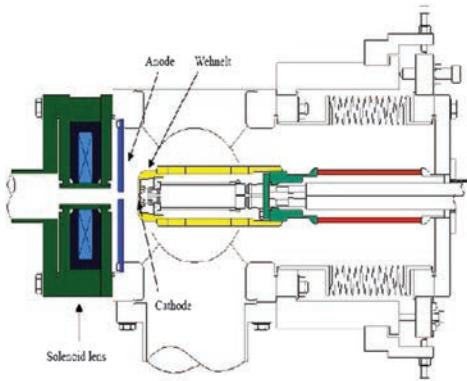


Figure 1: A DC gun incorporating a single  $\text{LaB}_6$  crystal thermionic cathode at the LNS-Tohoku facility in Japan[5] The beam exits to the left of the figure.

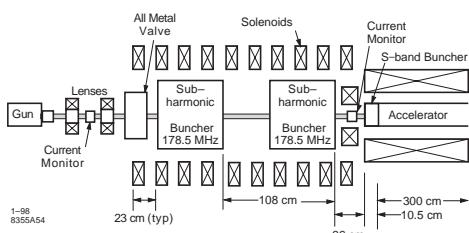


Figure 2: Schematic of the Stanford Linear Collider injector with DC HV gun and sub-harmonic prebunchers.

After a suitable drift, the tail of the bunch catches up with its head. In this way bunches are formed with widths of about  $10^\circ$  of the fundamental harmonic of the bunching cavities. This velocity modulation arrangement is referred to as a preinjector as it functions to prepare the beam for subsequent acceleration in a linear accelerator. An example of such a system is shown in Fig.2 which is a schematic of the injection system for the Stanford Linear Collider[1]. The DC Gun generates a 120 keV electron beam which is subsequently modulated by a 178.5 MHz standing wave rf buncher. This is followed by a 4 cell 2856 MHz traveling wave buncher. After a suitable drift, the beam is injected into a standard SLAC 3m 2856MHz traveling wave structure. This system is capable of producing beam bunches of  $5 \times 10^{10}$  electrons with a FWHM of 20ps.

Several simulation tools have been developed to aid in the design and evaluation of DC guns and

preinjectors. Among the most successful are: 1) EGUN[2], a 2D code capable of calculating beam profiles in a DC gun, including space-charge effects; 2) PARMELA[3], a  $2\frac{1}{2}$ D code which incorporates space-charge effects along with beam line elements including magnets and rf cavities; 3) SUPERFISH/POISON[4], a suite of 2D codes used for the calculation of both static and dynamic fields in magnets and rf cavities; MAFIA[5] (now Microwave Studio), a 3D particle-in-cell code which can be used when a full 3D simulation is required.

## References

- [1] H. Hanerfeld, et al, PAC85, p2510
- [2] W.B. Herrmannsfeldt, *EGUN, An Electron Optics and Gun Design Program* (1988) SLAC-331
- [3] Lloyd Young, James Billen, *THE PARTICLE TRACKING CODE PARMELA*, PAC03, (2003) p3521.
- [4] K. Halbach, R. F. Holsinger, PA 7 (1976) 213-222.
- [5] R.Klatt, et al, LINAC86, SLAC-303, 276 (1986)

### 7.1.1.5 Normal conducting rf photo guns

*D.H. Dowell, SLAC*

Normal Conducting RF (NCRF) photocathode guns produce high brightness beams and operate with average currents from microamperes to 10's of milliamperes [1]. Practical RF guns fall into the two general operational groups of low and high duty factor, or low and high average current. Low average current guns operate at high peak fields and produce very bright, high peak current beams. By necessity high average current rf guns operate at lower peak fields and generally produce beams with nearly the same transverse emittance as the high field guns, but at lower peak current and lower peak brightness. The peak cathode field is limited by electrical breakdown which is less likely at high RF frequencies [2] which also favors low duty factor operation. Typical peak fields are 100 to 200 MV/m for pulsed guns operating at a few GHz, while peak fields of 5 to 25 MV/m are common for milliampere to ampere capable guns operating in the frequency range of 100 to 900 MHz. High duty factor systems are constrained by thermal management of the gun's high average power dissipation [3] and RF breakdown which has a lower threshold for long RF pulses. As a rule of thumb, the surface power density should not exceed an average power of

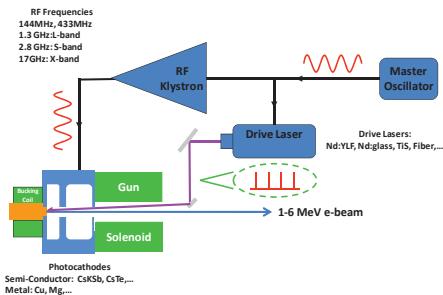


Figure 3: Parts of the RF photocathode gun system.

20 watts/cm<sup>2</sup>. The high power dissipation raises the cavity surface temperature which increases the outgassing rate and degrades the vacuum [4]. The higher vacuum significantly shortens the lifetime of a cathode already stressed by producing milliamperes to ampere currents. Pulsed guns typically are constrained by peak field and pulsed heating [5] rather than average power dissipation [19]. Convenient relations for the frequency scaling of the shunt impedance, power dissipation and other parameters in RF structures are given in Ref.[6].

Coupling the RF power into the cavities is an important design issue for all RF guns. Both pulsed and cw guns require careful attention to the heating around the RF ports in the cavity: pulse heating for pulsed guns and high power dissipation for CW guns [7]. Localized heating by the RF currents flowing around the RF ports produces mechanical stresses which distort the cavity's shape and detune the gun. Asymmetric or unbalanced power feeds also distort the axial RF field leading to additional emittance [8, 9]. Therefore, the gun fields are optimized using 3D codes [10] which include the effects of coupling ports, probes and other penetrations. Typically the power coupling ports are often the largest source of field asymmetry which can be greatly reduced by using either a coaxial RF feed [11] or a dual feed with a racetrack shape for the cell the power is coupled into [9]. Some innovative designs couple the power into the cavity through a coaxial cathode stalk [12].

In addition to carefully shaping the spatial fields it is also important that the gun RF oscillates at the desired single frequency mode of the gun. This is an especially important issue for pulsed, multi-cell guns with RF modes spaced closely in frequency. Adjacent modes can be excited by the

envelope of the RF waveform and then exchange energy or beat at their separation frequency [13]. Time-dependent RF modeling shows the modes should be widely separated to mitigate this problem [14]. The RF fields of the high gradient gun not only rapidly accelerate the beam but it also produce significant focusing effects. In a balanced gun, the RF lens is principally located at its exit and it is defocusing. The defocusing strength is large and time-dependent in a high field gun. For example, a cathode field of 120 MV/m the  $\frac{1}{2}$  cell s-band gun has a focal length of approximately -12 cm for a beam exiting on crest. In nearly all gun designs this strong defocusing is cancelled with a magnetic solenoid which is also used for emittance compensation. The solenoid's field is dominantly needed to cancel the strong defocusing of the gun which leads to an RF emittance even for perfectly linear transverse fields [17]. In addition the solenoid provides the focusing needed for emittance compensation [15] to match the beam to the first booster linac according the Ferrario condition [16]. Similar to the gun, the solenoid magnetic field needs to be of high quality. Beyond its careful design and fabrication the solenoid requires measurement of its magnetic properties to determine the size and orientation of the field multipoles. As a result of such measurements normal and skew correction quadrupoles were installed in the LCLS gun solenoid and found to be important for minimizing the emittance [20].

## References

- [1] D.H. Dowell, J.G. Power, "Summary of Working Group 4: High Average Power Facilities", UCLA High Average Power Workshop, Jan. 2009, To be published in AIP Conference Proceedings.
- [2] G.A. Lowe and J.W. Wang, SLAC-PUB-4647, May 1988
- [3] S.S. Kurennoy et al, Proc. of PAC2003, p. 920
- [4] D.H. Dowell et al, NIM A356 (1995) 167-176
- [5] D.P. Pritchau et al, PRST-AB 5 (2002) 112002
- [6] T. Wangler, RF Linear Accelerators, pp. 50
- [7] S.S. Kurennoy et al, PAC03, p. 3515
- [8] Z. Li et al, PAC05, p. 2176
- [9] L. Xiao et al, PAC05, p. 3432
- [10] Z. Li et al, "High Performance Computing in Accelerating Structure Design and Analysis", ICAP2004
- [11] F. Stephan et al, PRST AB 13 (2010) 20704 and references therein

## Sec.7.1: PARTICLE SOURCES

- [12] R.A. Rimmer, PAC05, pp. 3049-3051
- [13] J. Schmerge et al, "The 1.6 cell Gun Correlated Energy Spread Dependence on  $\pi$  and 0 Mode Amplitude," Proceedings of High Brightness Electron Beam Workshop, Erice Sicily, October 10-14, 2005
- [14] C. Limborg-Deprey et al, PAC05, p. 2233
- [15] L. Serafini J.B. Rosenzweig, PR E55, p. 7565
- [16] M. Ferrario et al, Proc. of the 2nd ICFA Adv. Acc. Workshop on "The Physics of High Brightness Beams", UCLA, Nov. 1999 & SLAC-PUB-8400
- [17] K.-J. Kim, NIM A275 (1989) 2001-218
- [18] D.H. Dowell et al, NIM A528 (2004) 316-320
- [19] D.H. Dowell et al,  
<http://www-bd.fnal.gov/icfabd/Newsletter46.pdf>  
 and references therein
- [20] R. Akre et al, PRST AB 11 (2008) 30703

### 7.1.1.6 Superconducting rf photo guns

*J. Sekutowicz, DESY*

Superconducting photo-injectors (SRF-guns) are radio-frequency (RF) electron sources exploiting the photoemission phenomenon. They are based on superconducting (sc) technology and can generate low emittance intense electron bunches in cw mode. SRF-guns were proposed in late 80's [1] for the high peak brilliance free electron laser facilities operating cw that were being planned. A generic SRF-gun consists of a photo-cathode emitting electrons when irradiated by optical laser light in an sc cavity for rapidly boosting these electrons to high energy. RF-guns, both normal- and superconducting, for high bunch charge, must operate at high accelerating gradients to suppress space charge dilution of the emittance. SRF-guns are hoped to be technically superior to those using normal conducting cavities because they dissipate orders of magnitude less power at very high gradients when they operate cw, but the integration of photo-cathode into the sc cavity is challenging. The cathodes are usually made of materials ( $Cs_2Te$ ,  $CsK_2Sb$ ...) having good quantum efficiency (QE), which are non sc, and have limited life time which requires the injectors using these cathodes to be equipped with complicated load-lock mechanisms for cathode exchange. For low current injectors, a sc Pb cathode, which unfortunately has moderate QE, can be an option avoiding the load-lock need [2]. Three types of sc cavities have been considered for the SRF-guns, 1.5-cell and 3.5-cell at 1.3 GHz, and 0.5-cell at

0.704 GHz. In general, cavities operating at lower frequency and having fewer cells are more suitable for higher beam currents. The number of cells in high current injectors is very often limited by the power capacity of input coupler. Additionally, fewer cells allow for the emittance growth compensation scheme, developed for and commonly used in normal conducting injectors, by a solenoid field. In the case of SRF injectors the magnetic field acts only on bunches exiting the cavity, because it cannot penetrate through the sc cavity wall. An alternative compensation method has been proposed for the 3.5-cell cavities [4]. Computer modeling has shown that the emittance growth can be compensated by TE-like monopole mode excited in the last cell. The magnetic field pattern of this mode is similar to a solenoid field. The method is not yet tested. Three approaches have been proposed to integrate the photo-cathode with the sc cavity preserving its original intrinsic quality factor and avoiding an RF-power leakage through the load-lock mechanism.

**Choke filters** Figure 4 (left) shows a 1.3 GHz 0.5-cell SRF-gun cavity, part of the mechanism for exchanging and precise positioning of cathodes, and a cell-like choke filter preventing leakage of the RF-power. Figure 4 (right) shows 704 MHz 0.5-cell cavity for a high current injector with another solution, a double quarter-wave (QW) choke. In either injector the cathode is exposed to high electric fields boosting the emitted particles. The main difficulty in the first choke is proper adjustment of its frequency. In the later one it is hard to process multipacting.

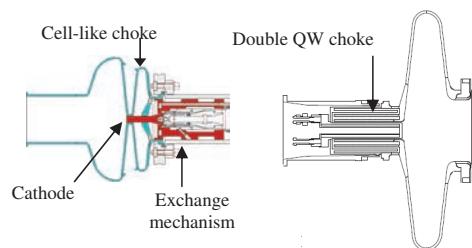


Figure 4: 0.5-cell cavities with cell-like choke (left) and quarter-like choke (right). Courtesy FZD and BNL.

**Pierce-gun with sc cavity** In this design, an SRF injector consists of the DC Pierce gun and a sc cavity (3.5-cell in Fig.5). The cathode is exposed to the DC electric field and not to RF field.

Table 2: Design Parameters of SRF injectors.

|                                     | Choke              |                     | DC-gun             |          | SC cathode |
|-------------------------------------|--------------------|---------------------|--------------------|----------|------------|
| f [GHz]                             | 1.3                | 0.704               | 1.3                | 1.3      |            |
| Cells                               | 3.5                | 0.5                 | 3.5                | 1.5      |            |
| E [MeV]                             | 9.5                | 2                   | 5                  | 6        |            |
| Cathode                             | Cs <sub>2</sub> Te | CsK <sub>2</sub> Sb | Cs <sub>2</sub> Te | Pb       |            |
| $\lambda$ [nm]                      | 262                | 532/355             | 266                | 213      |            |
| q [nC]                              | 0.077              | 1                   | 1.4                | 5        | < 0.06     |
| $\varepsilon$ [ $\mu\text{m rad}$ ] | 1                  | 2.5                 | 2                  | 2        | 1.2        |
| I <sub>beam</sub> [mA]              | 1                  | 0.5                 | 500                | 50       | 1-5        |
| f <sub>rep</sub> [MHz]              | 13                 | 0.5                 | 352                | 9.38     | < 81.25    |
| Status                              | Operation+R&D      |                     | R&D                | R&D      | R&D        |
| Institution                         | FZD                |                     | BNL                | IHEP PKU | DESY       |
| References                          | [3, 4]             |                     | [7, 8]             | [5, 6]   | [2, 9]     |

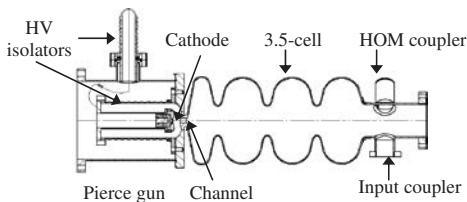


Figure 5: 3.5-cell cavity with attached Pierce-gun.  
Courtesy Peking University.

The emitted electrons are first accelerated in the DC field before they enter the accelerating sc cavity via small diameter channel. The channel is very deep in cutoff and thus prevents leakage of the RF-power into the DC section. The limited field on the cathode ( $< 10 \text{ MV/m}$ ) and drift of non relativistic bunches does not allow very low emittance ( $< 1 \mu \text{ mrad}$ ).

**Superconducting cathode** In the hybrid Pb-Nb gun the cathode, which is a spot of lead ( $\varnothing < 3 \text{ mm}$ ), is located in the center of the back wall of the sc cavity (1.5-cell in Fig.6). The hybrid gun [9] has potential to generate very low emittance beams with high charge/bunch but with low repetition rate. Moderate QE of the Pb cathode has to be compensated with more irradiating photons per

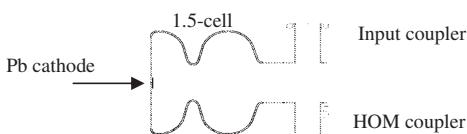


Figure 6: 1.5-cell with sc Pb cathode. Courtesy DESY/TJNAF.

pulse and higher energy ( $\sim 6 \text{ eV}$ ). Therefore the bunch repetition rate must be kept low ( $< 1 \text{ MHz}$ ) to avoid quenching of the sc cavity caused by the laser power.

One should also note that deposition of the superconducting cathode on the back wall is technologically demanding and the presence of the Pb spot makes cleaning of the sc surface difficult.

**Ultimate parameters and state-of-the-art** The design parameters for the SRF injectors discussed above are displayed in the table. They have been only partially demonstrated so far. Practically, these devices having the potential to generate high quality beams are still in the R&D stage. Up to now two of them at FZD and IHEP PKU delivered beams in 2002 [3] and 2003 [10] respectively. In January 2010 the SRF injector with 3.5-cell cavity at FZD was put in operation for the ELBE accelerator, even though the energy of generated particles it could deliver was below the specification for that injector. It delivers 400 pC per bunch at 3 MeV.

## References

- [1] A. Michalke et al, EPAC92 p1014
- [2] J. Smedley et al, PRST AB 11 (2008) 013502
- [3] D. Janssen et al, NIM A507 (2003) 314-317
- [4] A. Arnold et al, NIM 577 (2007) 440
- [5] R. Xiang et al, NIM A528 (2004) 321-325
- [6] F. Zhu et al, HEP and NP 31 (2007) 496-500
- [7] R. Calaga et al, SRF05 (2005) p506
- [8] I. Ben-Zvi, FEL2007, p293
- [9] J. Sekutowicz et al, TESLA-FEL R. 2005-09, DESY
- [10] J. Hao et al, NIMA 557 (2006) 138

### 7.1.1.7 DC photo guns

*B. Dunham, Cornell*

A DC photoemission electron gun consists of a photocathode mounted on a negative potential electrode, a grounded anode a short distance away, a laser, a high voltage insulator, and a high voltage power supply. The biggest advantage of any type of photoemission gun is that the electron bunches are generated by a laser synchronized to the appropriate RF accelerating frequency, thus most of the complex chopping and bunching hardware used with thermionic cathode guns can be dispensed with. Other advantages of DC photoemission guns are: any type of photo-emissive material can be used; it can be operated in DC mode, CW mode, or pulsed mode with arbitrary pulse widths and repetition patterns; has a simplified high voltage system since no filament heater or grid driver is needed; and can provide a wide range of peak and average power beams. Some disadvantages are: a laser is required; lower field gradient and beam energy than can be obtained with an RF gun; and bunch charges typically below a few hundred pC for high quality beams. An example of a photoemission gun that can be used with any type of cathode is shown in Figure 7 [1].

When the required beam voltage is 225 kV or less, industrial HV components are available. One can purchase a conical insulator, matching cable and high voltage power supply and quickly put together a photoemission gun [1]. Above 225 kV, ceramic bushings are used, either made from a solid insulating tube or stacked ceramic rings. A metal tube down the center supports the negative electrode and photocathode. The dimensions for the inner tube and insulator can be estimated from:

$$E = 1/R_{inner} \cdot V / \ln \left( \frac{R_{outer}}{R_{inner}} \right),$$

where  $V$  is the gun voltage,  $R_{inner}$  the support tube radius,  $R_{outer}$  the insulator radius and  $E$  the electric field on the inner tube. A good upper limit for  $E$  is 10 MV/m. For 750 kV, values of  $R_{inner} = 50$  mm and  $R_{outer} = 220$  mm result in a field of 10 MV/m on the inner conductor. For the height of an insulator, 250-300 mm in air is typically sufficient for 100 kV, or 250 kV in 1 bar SF<sub>6</sub>. For 500 kV, 500-600 mm is sufficient, depending on the gas pressure. The insulator material is most often alumina because of its strength and brazing properties. Plastics can be used when vacuum outgassing is not a concern.

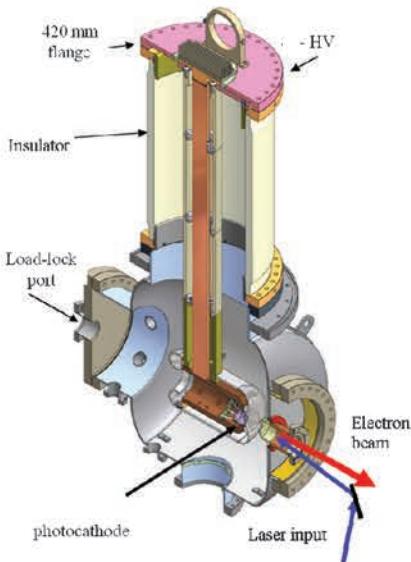


Figure 7: An example of a DC photoemission gun vacuum chamber, including a port for adding a load lock system for changing cathodes if required. The insulator and high voltage electrodes are mounted vertically so that the load-lock can be connected to the ground-side potential. For high voltages ( $>\sim 200$  kV), the insulator end is often enclosed in a SF<sub>6</sub> pressure vessel.

The insulator vacuum seal is made by brazing [2] a metal ring to it and then welding the ring to a flange. Figure 8 shows a particular joint that works well for the large insulators. Kovar has the closest expansion versus temperature match to alumina, and is often used for that reason.

The gun design shown in Figure 7 is prone to breakdown for voltages approaching -500 kV, which can occur during HV conditioning when field emission sites on the central tube emit electrons that hit the insulator. A different insulator design (Figure 9) can be used to block any electrons from reaching the insulator surface.

High voltage power supplies are commercially available up to -500 kV for currents below 10 mA. Above 10 mA isolated core transformers are used. A modern implementation is the Cross Transformer [3] which has extremely low capacitance, resulting in low stored energy and a fast response time. The insulator and HV power supply can be connected by a cable or by enclosing both in a pressure tank and connecting them with

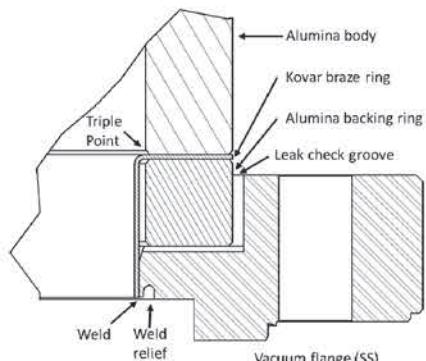


Figure 8: A ‘sandwich’ type braze joint. A kovo ring is brazed between the alumina body and a backing ring.

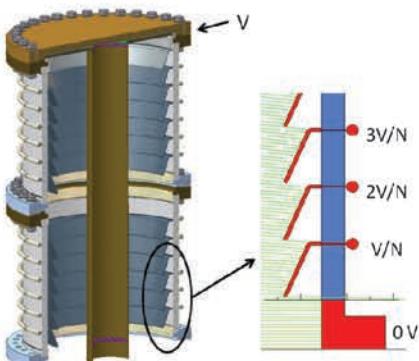


Figure 9: An insulator made up of alumina rings separated by braze rings designed to hold off a total voltage  $V$  (left). Internal chevrons block electrons field emitted from the central tube from reaching the insulator (right). Each of the  $N$  rings is held at a graded potential using external resistors. Electrons emitted from the tube are shown hitting the chevrons (right), thus protecting the insulator.

a tube or resistor. HV cables can be difficult to use and have high capacitance.  $SF_6$  is used as a dielectric gas in the tank, with 2.5X the voltage holdoff of air. For perfect parallel plates the  $SF_6$  strength is 89 kV/cm/bar, but must be derated for non-ideal geometries and safety factors. All HV devices must be conditioned above the desired voltage to minimize arcs during operations, typically 20-50% above. For sensitive cathodes, this is critically important as arcs can cause

vacuum spikes and damage the cathode. There are several methods for processing HV systems, including in-situ heating to remove absorbed gases, current conditioning, conditioning with multiple breakdowns, plasma processing, gas conditioning, and pulsed processing [4] A low capacitance HV power supply is beneficial in order to reduce the stored energy, and thus reduce the energy that can go into an arc ( $< 10$  Joules is a good upper limit) during processing. A resistor between the gun and power supply limits the current to the arc, with a value sufficient to keep the current below a few mA maximum at all times. Resistances of 50 to 200 M-Ohm are often used for voltages of hundreds of kV. The resistor must also be able to withstand the full voltage for the short duration of multiple arcs, and withstand the energy of the arcs.

## References

- [1] P.A. Adderley et al, PRST AB 13 (2010) 010101
- [2] F. Rosebury, Handbook of Electron Tube and Vacuum Techniques, AIP, Woodbury, NY, 1993
- [3] U. Uhmeyer, Proceedings of the 18th International Symposium on High Energy Spin Physics, AIP Conference Proceedings 1149, 1099 (2009)
- [4] M.E. Cunco, 18th International Symposium on Discharges and Electrical Insulation in Vacuum, 721-729, Eindhoven (1998).

### 7.1.1.8 Continuous duty preinjectors

*I. Bazarov, Cornell*

This class of devices serve as electron sources for linac-based accelerators, including recirculating linacs (e.g. CEBAF at TJNAF), ERLs (synchrotron radiation sources and electron-ion coolers), or continuous duty linacs operating without recirculation. The preinjector’s role is both to boost the energy of the electron beam as well as to prepare the phase space of the beam (both transverse and longitudinal) for injection into subsequent accelerating sections. Typical components of the preinjector are (following an electron gun): an RF bunching system, an energy booster, and magnetic optics for beam matching and transport.

**Superconducting vs. normal conducting** The continuous duty operation defines the choice of RF accelerating structures. For applications not requiring large energy per cavity cell (typically a fraction of MV), normal conducting structures represent a feasible solution. An example of

all normal conducting continuous duty preinjector is Novosibirsk FEL driver [2], operating with RF frequency of 180.4 MHz and bunch repetition of 22.5 MHz. Another example of continuously operating normal conducting RF is the buncher cavity that requires only a moderate voltage ( $\leq 200$  kV) [3]. At higher accelerating gradients, normal conducting cavities experience prohibitive Ohmic losses at cavity walls:  $P_c = V_c^2/(R/Q)_c Q_0$  for power dissipation in a cell, where  $V_c$  is the voltage per cavity cell,  $(R/Q)_c$  is the shunt impedance (typically between 100 to 200  $\Omega$ ), and the quality factor  $Q_0$  (on the order of  $10^4$  for copper) Therefore, superconducting RF with its 5 orders of magnitude higher  $Q_0$ 's is the technology of choice for applications requiring  $\geq 1$  MeV energy gain per cavity cell.

**High average current preinjectors** Design considerations vary significantly when high average current is required (10's of mA) as opposed to a lower current operation ( $\leq 1$  mA). As beam power reaches 10's of kW to MW levels any intercepting techniques for bunch manipulation such as RF deflection and chopping become problematic. Coupling large amounts of RF power into the beam drives the design of accelerating SRF structures: power transmission rating of RF windows limits the number of cells per cavity per input coupler, low loaded Q values require very strong coupling to the beam, so the care must be taken to avoid input couplers kicking the beam. This effect is usually mitigated by placing symmetrical input couplers [4]. Finally, the bunches in high average current preinjectors tend to be heavily space-charge limited requiring exquisite control of optics to achieve lowest transverse emittances [5].

**Cornell ERL injector** An example of a high average current continuous duty preinjector is Cornell ERL injector prototype [6], shown in Fig.10. Table 3 lists the main injector parameters.

The accelerating RF structures are superconducting with the exception of the RF bunching cavity. Five two-cell SRF cavities can each de-

liver 1 to 3 MeV to the beam [7]. The bunching RF cavity is a single cell copper structure [3] that operates at zero-crossing of RF phase. Assuming the longitudinal compression waist is to occur at distance  $L$  from the buncher (typically inside the first accelerating cavity), the buncher voltage, defined as energy gain per electron charge for a speed of light particle, is given by (ignoring the space charge):

$$eV_{bun} = \frac{\lambda_{RF}}{2\pi L} m_e c^2 \gamma (\gamma^2 - 1).$$

Here Lorentz factor  $\gamma$  corresponds to the energy at which the bunching takes place and  $\lambda_{RF} = c/f_{RF}$  is the RF wavelength.

**State of the art for high average current** The highest average current demonstrated from a continuous duty factor preinjector is 50 mA from Cornell ERL photoinjector in its commissioning stage in early 2012. The Boeing/LANL RF gun has demonstrated 32 mA average current at 25 per cent duty factor [1]. 30 mA has been achieved at Novosibirsk FEL [2]. TJNAF FEL photoinjector has delivered up to 9 mA of beam current [8]. *Technical challenges.* The challenges towards realizing  $\geq 100$  mA type device include gun performance (the combination of quantum efficiency and longevity for photocathodes used in photoemission guns [9]), control of space charge throughout the entire preinjector region, long range wakefields that can both degrade the beam performance and deposit unwanted HOM power into SRF cavities, as well as ion effects. Strong RF fields and their corresponding RF focusing may lead to emittance degradation. E.g., transverse emittance after a cavity is given by  $\varepsilon_n^2 = \varepsilon_0^2 + \varepsilon_{kick}^2 + \varepsilon_{foc}^2$ , where  $\varepsilon_0$  is the initial normalized emittance prior to entering the cavity, and RF kick and RF focusing induced emittances (see 7.1.1.5) are given respectively by  $\varepsilon_{kick} = \frac{1}{mec} \left| \frac{\partial p_x}{\partial z} \right| \sigma_x \sigma_z$  and  $\varepsilon_{foc} = \frac{1}{mec} \left| \frac{\partial^2 p_x}{\partial z \partial x} \right| \sigma_x^2 \sigma_z$ . Here  $\sigma_x$  and  $\sigma_z$  are transverse and longitudinal dimensions of the bunch, and the prefactors are derivatives of the transverse momentum delivered by the cavity to the beam with respect to transverse,  $x$ , and longitudinal,  $z$ , coordinates characterizing the RF kick,  $\frac{\partial p_x}{\partial z}$ , and RF phase dependent focusing,  $\frac{\partial^2 p_x}{\partial z \partial x}$ . For the symmetrical coupler configuration  $\frac{\partial p_x}{\partial z} = 0$  and only rf focusing term is important, which scales as  $\frac{\partial^2 p_x}{\partial z \partial x} \propto \frac{1}{\gamma}$ . This effect is particularly important for

Table 3: Injector design parameters.

|                             |           |
|-----------------------------|-----------|
| Beam energy                 | 5-15 MeV  |
| Charge per bunch            | 77 pC     |
| Max avg current             | 100 mA    |
| Rms bunch length            | 2 ps      |
| Emittance (normalized, rms) | 1 $\mu$ m |
| Operating frequency (CW)    | 1.3 GHz   |

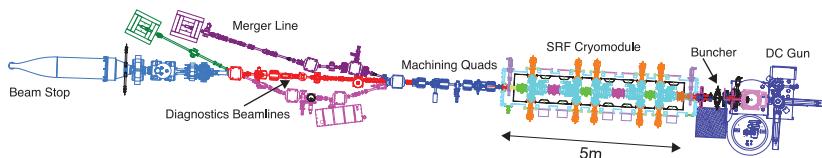


Figure 10: The layout of the Cornell prototype ERL photoinjector. The beam direction is to the left.

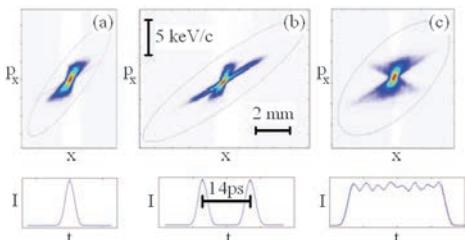


Figure 11: Effect of rf-focusing induced emittance growth – X-shape in (b) and ‘bow-tie’ shape in (c) - bottom row shows corresponding current vs. time.

low energy beams with large dimensions as illustrated in Fig.11, which shows the effect at 6MeV energy at Cornell ERL injector. The effect can be partially canceled through a particular choice of off-crest phases for individual rf cavities.

## References

- [1] D.H. Dowell et al, Appl Phys Let 63 (1993) 2035
- [2] N.A. Vinokurov et al, FEL 2009 Proceedings, p. 447
- [3] V. Veshcherevich, S. Belomestnykh, PAC03 p1198
- [4] V. Veshcherevich et al, PAC 03 p.1201
- [5] I.V. Bazarov, C.K. Sinclair, PRST AB 8 (2005) 034202
- [6] I.V. Bazarov et al, EPAC02 p644
- [7] V. Shemelin, PAC03 p2059
- [8] C. Hernandez-Garcia, PAC05 p3117
- [9] L. Cultrera et al, PRST-AB 14 (2011)120101

### 7.1.2 Positron Sources

#### 7.1.2.1 Tungsten targets S. Ecklund, SLAC

**Pair production** Positrons can be created by pair production or by beta decay of nuclei. Both are used as sources of positrons but pair production is the more common provider for accelerators, usually because of desired time struc-

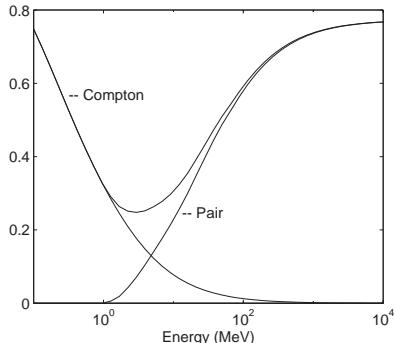


Figure 1: Probability per  $X_0$  of  $e^+e^-$  pairs and Compton scattering vs incident photon energy.

ture and high phase space density. A positron system consists of an incident beam hitting a target, followed by focusing devices to collect the positrons and an accelerator to provide the required energy.

The pair production cross section [1] is rapidly rising from threshold, becoming nearly constant at incident energies  $>$  a few hundred MeV.

In practice the positron production and the photon production are accomplished in a common target. A thick target results in a rapidly increasing multiplication of the flux of positrons until their energies fall too low to continue the cascade. For design of sources, a computer program, typically EGS [2] or GEANT [3], is used to calculate positron yields. These programs can deal with the complications of the design geometry and pair, Compton, and bremsstrahlung cross sections including effects of screening, multiple scattering, and ionization.

**Cascade showers** A considerable gain in numbers of  $e^\pm$  is realized by the cascade shower initiated by a  $e^-$  (or  $e^+$  or  $\gamma$ ) hitting a thick target. The longitudinal scale is set by the radiation length  $X_0$  of the material. The transverse properties are set by the transverse momentum, which is

## Sec.7.1: PARTICLE SOURCES

Table 1  $e^+$  yields for Cu and W

| E <sub>-</sub>                   | 17 GeV |     | 50 GeV |      |
|----------------------------------|--------|-----|--------|------|
|                                  | Cu     | W   | Cu     | W    |
| X <sub>0</sub> [mm]              | 14.3   | 3.5 | 14.3   | 3.5  |
| Y(6X <sub>0</sub> ) <sup>*</sup> | 2.8    | 5.8 | 7.4    | 13.4 |
| 2<E <sub>+</sub> <5 <sup>†</sup> |        |     |        |      |
| σ <sub>x</sub> [mm]              | 2.1    | 1.1 | 1.8    | 1.1  |
| σ <sub>px</sub> <sup>††</sup>    | 3.0    | 3.3 | 3.0    | 3.5  |

<sup>†</sup>MeV, <sup>††</sup>MeV/c, \*Yield / e at 6 rad. lengths

dominated by multiple scattering, and levels out at about 12 MeV/c because of the stopping of low energy  $e^+$ . The energy spectrum of the positrons is peaked at the few MeV level and falls inversely with energy. The cascade properties lead to a maximum positron yield almost linear with incident energy occurring at a depth into the target which increases as  $\log E$ . High Z targets are chosen for higher yield and smaller transverse size as illustrated in Tab.1.

Numerical fits [4] to EGS Monte Carlo calculations for W-Re targets with  $E_-$  from 0.3 to 30 GeV follow in Eqs.(1-2):

$$z_{\max} = 1.1 \ln(E_-) + 3.9 \quad (1)$$

where  $E_-$  is in GeV and  $z$  in  $X_0$  units. The peak yield value is parametrized as

$$\frac{1}{N_-} \frac{dN_+}{dE_+} = \frac{E_-(0.57 - 0.056 \ln E_-)}{E_+^{1.75}} \frac{P_r^2}{P_r + 2.5} \quad (2)$$

where  $P_r$  is the maximum  $e^+$  transverse momentum accepted, MeV/c and  $E_+$  the  $e^+$  energy in MeV,  $E_-$  the  $e^-$  energy in GeV. (See [17] for a comparison of transverse momentum distributions from thick and thin targets.)

### Material strength issues

Pulse Heating and Stress: The process of ionization ( $dE/dx = 2 \text{ MeV/gm/cm}^2$ ) produces target heating.

$$\Delta T = (2N_+/\pi r^2 c_p) dE/dx \quad (3)$$

where  $c_p$  is the target specific heat,  $r$  the beam radius. The stress following from this can be complicated but is of order

$$\sigma = \alpha_T E \Delta T / [2(1 - \mu)] \quad (4)$$

where  $\alpha_T$  is the thermal expansion coefficient,  $E$  Young's modulus and  $\mu$  Poisson's ratio (0.25-0.3) for the target material. Besides remaining

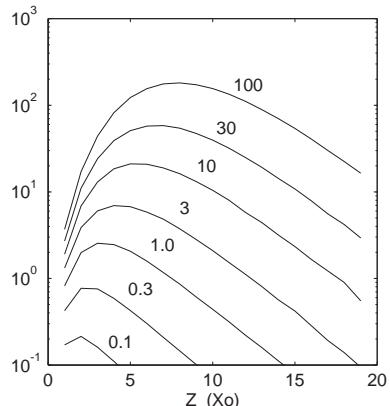


Figure 2: Total yield ( $N_+/N_-$ ) in tungsten vs.  $Z$ , for incident energies between 0.1 and 100 GeV.

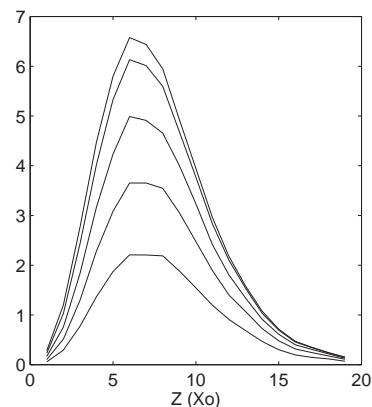


Figure 3: Collected yields in W for  $E_- = 30$  GeV, and maximum  $E_+$  of 5, 10, 20, 50, 100 MeV, within a 4-D phase ellipse ( $x, p_x, y, p_y$ ) with 2 mm and 2.5 MeV/c intercepts.

below the ultimate tensile strength one needs to take fatigue and expected cycle life into account in choosing a safety factor. Beam tests of W and Ta targets at SLAC [5] at 20 GeV indicate failure above incident beam densities of  $10^{11} \text{ e/mm}^2$  per linac pulse. Accurate calculations would involve EGS for yield and heating, the latter being input to a 2- or 3-D mechanical engineering code to find temperatures and stresses vs time in the proposed geometry [5].

Average Power Capabilities: Pulse heating and repetition rate together with heat conductivity of the target and heat transfer to the cooling

medium determine target size and need for a moving target. Typically moving targets are needed above several kW power absorption and are several cm in radius. Up to a meter in radius is needed for designs at the MW level [6]. Liquid targets [7] hold promise for high power densities but pose problems of containment or splashing if used in an open stream.

**Collection systems** Angular spreads  $\sim$  radian, spatial extents of  $\sim 1$  mm and large energy spread of the  $e^+$  from the target present challenges for collection; matching requires beta of 1 mm with large energy acceptance. Typically only a fraction of the positrons can be collected. Together with the accompanying  $e^-$  they also suffer from space charge and wakefield effects.

**Transverse collection:** Most commonly a uniform solenoid [8, 9, 10] with field parallel or antiparallel to the beam direction is used. It is wide band with focusing dependent on  $p_r$  only, and not on  $p_{\parallel}$ . The admittance of a channel of radius  $r_a$  is

$$\epsilon \left[ \frac{\text{Gev}}{c} \cdot \text{m} \right] = r_a p_r = 0.3 \frac{B[\text{T}] r_a^2 [\text{m}^2]}{2} \quad (5)$$

where Area (phase space)  $= \pi\epsilon$ . Most collection systems have several meters of solenoid before switching to quadrupole FODO channel when the energy spread has decreased to  $\sim 10\%$ .

To match the uniform solenoid to the beam at the target, an optical transformer is needed. A slowly varying longitudinal magnetic field is the common solution;  $B(z) = B_0 / (1 + \alpha z)$ . The field is adiabatic if  $\alpha \ll 0.3 B/p$  (units as above): higher momentum requires longer solenoid. In an adiabatic solenoid with  $B$  varying from  $B_0$  to  $B_f$  the transverse momentum is compressed by the factor  $\sqrt{B_f/B_0}$  and expanding the size by the inverse factor.

Another approach is the quarter wave transformer. A solenoid field,  $B_l$  of length  $L$  produces a focusing  $k = B_l/2(B\rho)_+$  and is a quarter wave transformer if  $kL = \pi/2$ . In front of a uniform solenoid of field  $B_f$  such a quarter wave transformer gives the transverse transformation factor  $B_f/B_l$  and the inverse for the momentum.

In yet another approach a flux concentrator [11] has been used to produce fields as large as 6T near the target. In it eddy currents channel flux into a small volume thereby magnifying the field. These fall somewhere between adiabatic and quarter wave.

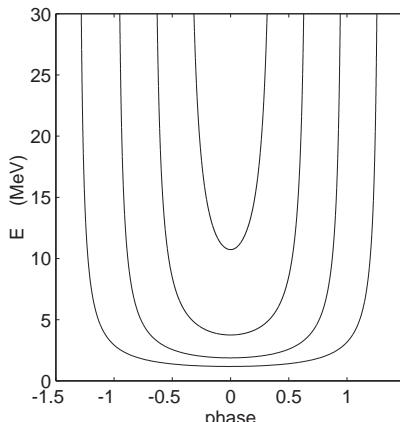


Figure 4: Longitudinal phase trajectories.

Currents along the beam might also be used as a matching lens. A lithium lens has been considered [12].

**Longitudinal collection:** Positrons of energy  $E$  and  $p_r/p \ll 1$  moving in an accelerator with rf wavelength  $\lambda$ , gradient  $G$ , and phase velocity  $\beta_p$ , move on paths given by

$$E/\beta_p - \sqrt{E^2 - m^2} = \lambda G (\cos \phi - \cos \phi_f) \quad (6)$$

shown in Fig.4 ([18]).

Particles move towards negative phase and approach  $\phi_f$  asymptotically. Matching at the beginning of the capture process can be enhanced by introducing a small drift space thereby producing an  $E\phi$  correlation matching the slope of the phase trajectories in the capture section downstream of the target. In designing to match the ultimate downstream device, e.g. a storage ring [13], one must take into account the transverse motion and space charge effects. The latter is complicated by the fact that as the electrons and positrons separate, the space charge neutralization disappears and the defocusing effect turns on. Various computer codes are used to calculate this as well as the optics of the transverse and longitudinal transport [14]. When space charge is significant, one gains by using a large  $G$  in the capture section(s) [15].

### Existing and proposed sources

Tab.2 describes the SLC and proposed NLC sources, both with W75Re25 targets [16].

## Sec.7.1: PARTICLE SOURCES

Table 2 Existing and Proposed Sources

|  | SLC     | NLC   |
|--|---------|-------|
| <b>General</b>                             |         |       |
| Ne <sup>+</sup> /pulse [10 <sup>10</sup> ] | 3-5     | 63    |
| Nbunch/pulse                               | 1       | 90    |
| $\tau$ pulse [ $\mu$ s]                    | 3ps     | 0.126 |
| bun. sp. [ns]                              | -       | 1.4   |
| rep. rate [Hz]                             | 120     | 180   |
| e <sup>-</sup> beam                        |         |       |
| E [GeV]                                    | 30      | 3     |
| Ne <sup>+</sup> /pulse[10 <sup>10</sup> ]  | 3-5     | 135   |
| Power [kW]                                 | 17-29   | 121   |
| <b>Target</b>                              |         |       |
| Thickness [X <sub>0</sub> ]                | 6       | 5     |
| $\sigma_r$ [mm]                            | 0.6     | 1.6   |
| $\Delta T$ [K]                             | 200-300 | 200   |
| Power deposit[kW]                          | 4.2-6.0 | 40    |
| Ne <sup>+</sup> out [10 <sup>10</sup> ]    | 180-300 | 800   |
| <b>Collection</b>                          |         |       |
| Initial B [T]                              | 7       | 7     |
| End field [T]                              | 0.5     | 0.7   |
| Taper length[m]                            | 0.15    | 0.2   |
| $\lambda_{rf}$ [m]                         | 0.105   | 0.21  |
| min. iris [mm]                             | 9       | 20    |
| G [MV/m]                                   | 30      | 20    |
| Ne <sup>+</sup>                            | 4.5-7.5 | 126   |
| Admittance [ $\gamma x'x$ ,m]              | 0.01    | 0.06  |
| Energy                                     | 20      | 30    |
| Acceptance [MeV]                           |         |       |

## References

- [1] B. Rossi, High-Energy Particles, Prentice-Hall (1956)
- [2] W. Nelson, H. Hirayama, D. Rogers, SLAC-265 (1985)
- [3] R. Brun et al, CERN-DD/78/2 Rev. (1978)
- [4] H. Braun, SLAC-CN-388 (1991); H Braun et al, EPAC 92, p.530
- [5] J.A. Hodgeson, E.M. Reuter, PAC 91, p.1996; E.M. Reuter et al, PAC 91, p.1999; S. Ecklund, SLAC-CN-128 (1981)
- [6] P. Sievers, M. Hofert, CERN CLIC Note-66 (1988)
- [7] G.I. Sylvestrov (BINP), Proc. Workshop on New Positron Sources for Linear Colliders, SLAC-R-502 (1997)
- [8] a) R.H. Helm, SLAC-4 (1962); R.H. Helm, R. Miller, in Linear Accelerators, Section on Particle Dynamics, North Holland (1970)
- [9] R.B. Neal, ed, The Stanford Two-Mile Linear Accelerator, Benjamin (1968)
- [10] R. Chehab, LAL/RT 92-17
- [11] A.V. Kulikov, S.D. Ecklund, E.M. Reuter, PAC 91, p.2005
- [12] B.F. Bayanov et al, EPAC 88, p.263
- [13] J.E. Clendenin et al, 14th Int. Conf. on High Energy Part. Acc. (1989); PA 30 (1990) 105
- [14] R. Pitthan et al, PAC 91, p.2098
- [15] J.E. Clendenin, S.D. Ecklund, H.A. Hoag, PA 30 (1990) 85
- [16] Proc. Sources 94, Int. Workshop on e<sup>+</sup>e<sup>-</sup> Sources and Pre-Accelerators for Linear Colliders (1994) p.234
- [17] K. Flöttmann in Handbook of Accelerator Physics and Engineering 1st ed, 3rd print (2006) p.479
- [18] R.Helm, R.Miller in P.Lapastolle, A.Septiereds, Linear Accelerators North Holland (1970) p.115

### 7.1.2.2 Conversion of undulator radiation

*A. Mikhailichenko, Cornell*

**Overview** Positron sources based on conversion of high energy photons are being developed for future linear colliders [1]–[4] and is now the baseline for ILC [5]. The rate of positron production required for ILC is  $\sim 3.9 \cdot 10^{14}$  e<sup>+</sup>/sec ( $2 \cdot 10^{10} \cdot 2625 \cdot 5$ [Hz]; with 50% overhead). In the initial configuration the polarization will be  $\sim 35\%$  and can be improved to  $\sim 65\%$  in the future. The method of positron production envisioned is a two stage process. In the first stage circularly-polarized photons are generated in a helical electromagnetic field. These photons are then converted into positrons and electrons in a thin target. Longitudinal polarization is transferred from the primary photon beam to the secondary particles in accordance with the particle's energy. The scheme is shown schematically in Fig.5. Polarized electrons can be produced with this scheme as well. Primary electrons (or positrons) with energy  $\geq 150$  GeV are directed into a helical undulator having length  $< 200$  m. Perturbations of emittance and energy spread in the undulator are small thus this conversion system can be installed before the IP. For the initial ILC run when polarization is not an issue, the length of undulator could be cut to  $\sim 30$  m. The photons radiated have the energy of first harmonic i.e. 10–20 MeV. During passage through the undulator the beam loses 1–3 GeV depending on the conversion scenario (energy of primary beam, undulator length and  $K$  factor; (see Sec. 3.1.5)). The gammas reach the target after drifting  $\geq 150$  m in an evacuated tube, while primary electrons (or positrons) continue on their way to IP. The divergence of gamma flux  $\sim K/\gamma$  is bigger than the angular spread in the ILC beam. The collimator in front of target enhances the fractional polarization of the gamma flux. The target is thin,  $\sim 0.4 \cdot X_0$ . Immediately after the target an Optical

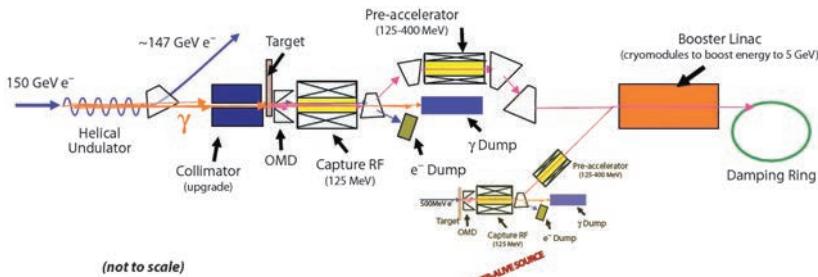


Figure 5: Conversion system of ILC [5]. Helical undulator is  $\sim 170$  m long. Distance between undulator and target is about the same.

Matching Device (OMD in Fig.5) is installed(See Sec. 7.1.2.1). The electron-positron mixture enters the capture RF structure and positrons, are separated from the remaining electrons, accelerated in a pre-accelerator and go to the damping ring ( $\sim 5$  GeV). There positrons will be cooled by synchrotron radiation before ejection towards the main linac entrance. The keep-alive source, shown in Fig.5, serves for filling the positron damping ring in absence of the main 150 GeV beam.

**Photon generation** For generation of photons, undulators using EM waves, helical static magnetic fields and helical crystals have been considered [1] Laser radiation, as a specific type of EM wave, was first considered in [6]. It was found, that the helical undulator with (quasi)static magnetic field is the only practical approach at present.

The photon energy for pair creation should be  $E_\gamma \geq 2mc^2$ . The cut-off energy of the first undulator harmonic is a Doppler-shifted spatial frequency of the undulator field

$$\begin{aligned} E_{\gamma \text{ max}} &\cong \hbar \omega_c = \frac{\hbar \Omega}{1 - \vec{n} \vec{v}} \Big|_{\vartheta=0} \\ &\cong \frac{2\hbar \Omega \gamma^2}{(1 + K^2 + \gamma^2 \vartheta^2)} \Big|_{\vartheta=0} \\ &\cong \frac{23.7 \cdot (E_e/50\text{GeV})^2}{\lambda_u [\text{mm}] (1 + K^2)} [\text{MeV}], \end{aligned}$$

where  $\Omega = 2\pi c/\lambda_u$ ,  $\vec{n}$  stands for unit vector in direction to observer,  $\vec{v}$  is the incident particle's velocity at the moment of radiation,  $E_e$  is energy of primary beam and  $\lambda_u$  is the undulator period. One can see that energy of quanta drops  $\sim 2$  times for angles  $\vartheta \sim 1/\gamma$ .

For,  $K \leq 1$ , the total number of quanta radiated by an electron can be approximated as

$$N_\gamma \cong 4\pi \alpha \frac{L}{\lambda_u} \frac{K^2}{1 + K^2},$$

where  $L$  is the undulator length and  $\alpha$  is the fine structure constant. The degree of photon circular polarization,  $\xi_2$  for  $K \leq 1$  can be expanded as  $\xi_2 \approx (1 - \gamma^4 \vartheta^4)/(1 + \gamma^4 \vartheta^4)$ , so for  $\vartheta \cong 1/\gamma$  the polarization comes to zero [10]. Thus a collimator cutting out the large angle radiation is needed for producing highly polarized photons.

#### Polarized positron generation

As seen in 7.1.2.1, higher photon energies are favored for pair production yield, e.g. the yield for 10 MeV photons exceeds that for 6 MeV by  $\sim 2$  x.

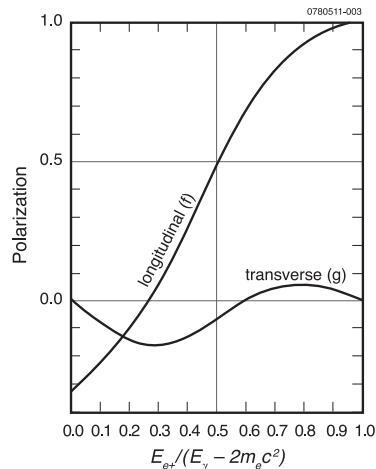


Figure 6: Polarization as function of fractional energy [8].

The polarization of created positrons as a function of its energy  $E(e^+)$  can be expressed as [8]

$$\vec{\zeta} = \xi_2 \cdot [f(E_+/E_\gamma) \cdot \vec{n}_\parallel + g(E_+/E_\gamma) \cdot \vec{n}_\perp] = \vec{\zeta}_\parallel + \vec{\zeta}_\perp$$

where functions  $f$  and  $g$  describe longitudinal and transverse polarization, shown in Fig.6,  $\vec{n}_\parallel$  is the unit vector along the initial direction of the gamma radiation,  $\vec{n}_\perp$  is the unit vector normal to it. As can be seen, the overall process is complex enough that numerical calculations are best done by simulation methods.

**Simulation of conversion** The expression for spectral density of the radiation for an undulator having length  $L$  has a form of sum over harmonic number [9, 10]

$$\frac{dN_\gamma}{dE_\gamma} = \sum_n \frac{dN_{\gamma n}}{dE_\gamma} = \frac{\alpha K^2 L}{\hbar c 2\gamma^2} \sum_{n=1}^{\infty} F_n(K, s),$$

where,  $F_n(K, s) = J_n'^2(n\kappa) + \frac{1+K^2}{4K^2} \frac{(2s-1)^2}{s(1-s)} J_n^2(n\kappa)$ ,  $\kappa = 2K\sqrt{s(1-s)/(1+K^2)}$ ,  $J_n$  is the Bessel function of the first kind and  $s = E_{\gamma n}/E_{\gamma nmax}$ . The number of positrons generated by a single photon in the target layer having thickness  $d\tau$  at the depth  $\tau$  counted from the entrance becomes

$$\frac{dN_+}{dE_+ d\tau} \cong 0.4 \frac{\alpha K^2 L}{\gamma^2 \hbar c} \frac{7}{9} (1 - E_{e^+}/E_{\gamma 1}) (1 - e^{-7\tau/9})$$

For  $E_0 = 150 \text{ GeV}$ ,  $L = 150 \text{ m}$ ,  $K^2 = 0.1$ ,  $\tau \cong 0.5$  (in units of  $X_0$ ) it gives  $(1/N_{tot})dN_+/dE_+ \cong 0.2 [1/\text{MeV}]$  where  $N^+$  is the integral of the equation over the positron energies and target thickness.

### Numerical simulation with Monte-Carlo

Several numerical codes have been used for modeling of positron conversion such as KONN [9], utilizing Monte-Carlo simulation of radiation from the undulator up to 10<sup>th</sup> harmonic, conversion in target, collection by Li lens, energy selection and further acceleration. Other approaches utilize EGS4, Geant4 and Fluka [11, 12].

**Energy deposition in target** The gammas create electron/positron pairs and Compton electrons; the ratio of  $\sigma_{\text{Compton}}/\sigma_{\text{pair}} \approx 1/(\gamma Z\alpha) \sim 10\%$ . In the energy range of interest the emerging particles lose energy at  $\sim 2 \text{ [MeV/g/cm}^2]$ .

To avoid target destruction by overheating, it is made as a rim spinning with angular speed  $\sim 2\pi 50 \text{ Hz}$  and diameter  $\sim 1 \text{ m}$ . The linear velocity of the rim at radius where the gamma beam

strikes reaches  $\sim 150 \text{ m/sec}$ . If energy  $Q$  is deposited within volume having mass  $m$ , we can estimate  $\Delta T[0 \text{ K}] \cong 2/c_V \cong 2A/25[\text{mol/g/oK}]$  where A is the Atomic weight. For Titanium, Tungsten and Lead the ratio of temperatures is  $\Delta T_{Ti} : \Delta T_W : \Delta T_{Pb} \cong 1 : 3.7 : 3.8$  suggesting the use of Titanium as the target material [4]. Use of Mercury may be advantageous for its low boiling temperature ( $\sim 357^\circ\text{C}$ ), so that the latent heat of vaporization ( $\sim 59.1 \text{ kJ/mol}$ ) plays an important role in the cooling process [9].

### Collection optics

As the angular spread of positrons is not as big as in the conventional method due to small thickness of target, collection optics (OMD) adds  $\sim 50\%$  to the total output. Use of collection optics has a peculiarity here as the spinning target rim perturbs the external magnetic field as result of eddy currents in the moving metal. The collection optics must be field-free at the target. There are two major candidates for OMD: the Lithium lens [9] (see section 7.2.12) and Flux concentrator [13].

### Undulators for conversion system

Several types of undulator have been considered for the conversion system of the linear collider. An undulator with SC windings was found to be best matched to the ILC positron source. SC undulators with period  $\sim 11 \text{ mm}$ ,  $K \sim 0.9$  have been developed and tested [2], [5].

**Experimental Tests of method** For experimental test of the undulator-based scheme of positron production, E-166 experiment was performed at SLAC [14]. The period of 1m long undulator manufactured was 2.54 mm, which for 46.6 GeV-beam and  $K = 0.17$  gives for the photon cut off energy  $\sim 7.89 \text{ MeV}$ . A polarization of  $\sim 85\%$  was confirmed. The Intensity obtained was in good agreement with calculations.

The scheme with Laser radiation is under development at KEK [15]. Latest experiments with laser back scattering [16] indicate that the rate  $\sim 10^8 \text{ e}^+/\text{sec}$  is achieved. It is below required  $10^{14} \text{ e}^+/\text{sec}$ , but the work is in progress. The approach using channeling radiation has also been explored [17].

### References

- [1] V.E. Balakin, A.A. Mikhailichenko, INP 79-85, Budker Institute for Nuclear Physics (BINP), Novosibirsk, 1979

Table 1: Possible conversion targets (all parameters are functions of temperature).

| Material        | Z  | $l_{X0}$ , cm | $\rho$ , g/cm <sup>3</sup> | $cv$ , J/g/K | $k$ , W/cm/K | $\alpha \cdot 10^6$ , 1/K | Melt T, °C | Y, GPa |
|-----------------|----|---------------|----------------------------|--------------|--------------|---------------------------|------------|--------|
| W-26Re          | 74 | 0.35          | 19.3                       | 0.13         | 1.3          | 5.0                       | 3422.      | 400.   |
| Ti              | 22 | 3.58          | 4.54                       | 0.52         | 0.22         | 8.4                       | 1668.      | 116.   |
| Ti-13V-11Cr-3Al | 22 | 3.56          | 4.84                       | 0.51         | 0.1          | 9.4                       | 1554.      | 101.   |
| Li              | 3  | 152.1         | 0.53                       | 0.28         | 0.84         | 4.6                       | 180.54     | 4.9    |
| Pb              | 82 | 0.56          | 11.3                       | 0.13         | 0.35         | 28.9                      | 327.4      | 16.    |

- [2] A.A. Mikhailichenko, PhD Thesis, BINP, 1986; (translated in CBN 02-13, Cornell, LEPP, 2002)
- [3] K. Flöttman, PhD Thesis DESY, 93-161 (1993)
- [4] K. Flöttman, J.Rößbach, DESY-M-91-11, Oct 1991, 19pp
- [5] J.A. Clarke et al, EPAC08-WEOBG03, Jun 25, 2008, 3pp
- [6] E.G. Bessonov, 15th Int. Conference on High Energy Accelerators, Hamburg, 1992, p. 138
- [7] W. Heitler, “The Quantum Theory of Radiation”, Oxford University Press, NY, 1954
- [8] H. Olsen, L.C. Maximon, Phys. Rev. 114, N3 (1959) 887
- [9] A.A. Mikhailichenko, (PAC 07) pp1974-1978
- [10] E.A. Bessonov, A.A. Mikhailichenko, Budker INP-1992-43, Jun 1992, 26pp.  
<http://ccdb4fs.kek.jp/cgi-bin/img/allpdf?199302032>
- [11] W. Gai, W. Liu, Int. Workshop on Positrons at Jefferson Lab, AIP Conf. Proc. 1160:67-73, 2009
- [12] F. Zhou, et al., SLAC-PUB-12239, Jan 2007
- [13] J. Gronberg, T. Piggott, et al., LLNL, a Talk at LCWS10 (2010) 27pp
- [14] G. Alexander et al, NIM A610 (2009) 451-487
- [15] T. Okugi et al, Jpn. J. Appl. Phys. 35 (1996) 3677
- [16] S. Miyoshi et al, e-Print: arXiv: 1002.3462
- [17] R. Chehab et al, PL B 525 (2002) 41

### 7.1.3 Polarized Protons and Heavy Ions

T.B. Clegg, U.N.Carolina-Chapel Hill  
W. Haeberli, U. Wisconsin-Madison

**Overview** Production of intense beams of polarized protons (or polarized H<sup>-</sup> ions) for injection into accelerators is based on the following steps: (i) production of a beam of hydrogen atoms polarized in electron spin; (ii) transfer of the electron polarization of the H atom to the proton; and (iii) ionization of the atoms by electron bombardment or charge exchange with other (unpolarized) atoms or ions. A review [1] and recent conference proceedings [2, 3] are informative. Requirements for ion sources are specific to individual accelerators (e.g. duty factor, phase-space acceptance). Polarized high-energy antiprotons arise from parity-violating decay of hyperons.

### Polarized ion sources based on atomic beams

[2, 3] Atomic-beam sources (ABS) produce a beam of thermal H atoms by dissociation of H<sub>2</sub> in an rf discharge. This beam is directed through differentially pumped vacuum chambers into multiple-element sextupole magnets [4] (usually rare-earth permanent magnets with  $B(poletip) \sim 1.6$  T; aperture = 0.8 to 3 cm) where the inhomogeneous field ( $B \sim kr^2$ ) focuses only atoms with electron spin parallel to  $B$ . These systems resemble closely those designed for polarized gas jet targets [5]. Subsequently, atoms in the electron-spin-polarized beam undergo adiabatic-fast-passage transitions [6, 7, 8] which transfer electron polarization to nuclear polarization  $P > 90\%$ , reversible in sign at up to 100 Hz. Ionization to produce H<sup>+</sup> or H<sup>-</sup> occurs by bombardment with electrons [9, 10] of  $\sim 100$  eV ( $H + e \rightarrow H^+ + 2e$ ) or by collisions with atoms ( $H + Cs \rightarrow H^- + Cs^+$ ) or ions ( $H + D^\pm \rightarrow H^\pm + D$ ) of energy chosen to optimize charge transfer [11]. Polarized H<sup>+</sup> output currents are 50 to 500  $\mu$ A (dc) within 1.2  $\pi$  mm-mrad, or 1 to 10 mA (pulsed at 0.2 ms with  $2 \times 10^{-3}$  duty factor) within 11  $\pi$  mm-mrad (normalized emittances, 80% of beam); typically  $P = 0.6$  to 0.8 is available. For H<sup>-</sup>,  $P = 0.6$  to 0.9, but beam currents are  $\sim 10$  times smaller than for H<sup>+</sup>; except 4.0 mA pulsed was achieved [12].

### Polarized ion sources based on optical pumping (OPPIS) [2, 13]

In these sources [14] unpolarized H<sup>+</sup> ions pick up a polarized electron from optically pumped alkali vapor (Na, or more typically Rb) inside a strong magnetic field to produce electron-spin-polarized atoms [13]. A very strong field ( $B \sim 20$  T) is used to maintain maximal atomic polarization [14, 15, 16]. Emittance degradation of the H<sup>+</sup> beam arising from entry into the strong field can be avoided [13] by double charge exchange so that the beam enters the  $B$ -field while neutral ( $H^+ + H_2 \rightarrow H + H_2^+$  prior to entry into the field;  $H + He \rightarrow H^+ + He + e$  prior to Rb cell). The electron-spin-polarized H

atoms emerge from the Rb cell into a region of sudden axial  $B$ -field reversal, where Sona transitions [17, 18] transfer the electron polarization to the nucleus. Subsequent ionization by charge exchange is used ( $H + Na \rightarrow H^- + Na^+$ , or  $H + He \rightarrow H^+ + He + e$ ) to produce the beam desired. These sources can produce dc polarized  $H^-$  (or  $H^+$ ) currents up to 1.6mA (or 1.3 mA) within  $2\pi$  mm-mrad (80% of beam), with  $P > 0.88$  [19]. Pulsed  $H^\pm$  beam currents over 10 mA have been reported [13].

### Polarized-ion sources based on the Lamb-shift

[20] Unpolarized beams of excited, metastable  $H(2S)$  atoms are produced by charge-exchange of unpolarized  $H^+$  ions in Cs vapor. Subsequently, proper configuration of rf and dc electromagnetic fields causes atoms in selected 2S hyperfine states to decay to the 1S ground state leaving the remaining 2S atoms nuclear polarized. They can then be ionized preferentially over the unpolarized 1S atoms by charge exchange in argon. Lamb-shift sources produce output polarized  $H^-$  beam intensities  $\leq 1\mu A$  and are no longer competitive. Techniques based on the Lamb-shift underlie the design of convenient, efficient monitors of beam polarization of hydrogen-like ions [21].

### Polarized-ion sources for deuterium and heavier ions

Deuterons: The above methods apply to deuterium as well. For the ABS, the polarization can be manipulated with adiabatic rf transitions to obtain pure vector polarization or pure tensor polarization (alignment) of reversible sign [7, 8], while OPPIS sources are limited to Sona transitions, which provide much less flexibility [22]. The Lamb-shift source permits selection of individual hyperfine states. Intensities of deuterium beams are similar to proton beams.

Heavier ions: Polarized beams of Li and Na are produced by the atomic-beam method, enhanced by optical pumping. ( $Li^+, Na^+, \sim 30\ \mu A$ ;  $Li^-, Na^-, 1\mu A$ ) [23, 24]. For methods of producing polarized  $^3He$  beams and polarized radioactive species, see [2, 25].

**Secondary beams of polarized protons and antiprotons** In the earliest polarized-proton experiments, the polarized particles were produced by scattering a primary beam from a (unpolarized) target, the polarization arising as a result of the strong spin-orbit force in hadronic interactions. For most applications this method has been superseded by the much more intense beams available from accelerators equipped with polarized-

ion sources. However, in high-energy accelerators, spin-precessors (Siberian Snakes) [26] are required to avoid depolarizing resonances (See also Sections 2.6.4 and 7.2.18). To avoid such precessors, at the FNAL polarized-beam facility a polarized-proton and -antiproton beam with momentum 185 GeV/c is obtained from the decay of lambda hyperons which in turn are produced by bombardment of a beryllium target with unpolarized 800 GeV/c protons. About  $3 \times 10^6$  protons ( $P > 45\%$ ) are produced per 20 s spill [27]. A dedicated new facility to produce stored polarized anti-proton beams by spin-filtering using a polarized H gas target has been proposed for GSI [28].

## References

- [1] W. Haeberli, Annu. Rev. Nucl. Sci. 17 (1967) 373
- [2] Polarized Sources, Targets, and Polarimetry, Proceedings of the 12<sup>th</sup> International Workshop, eds. A. Kponou, Y. Makdisi, and A. Zelenski, AIP Conf. Proc. 980 (2008)
- [3] Proc. 18th International Spin Physics Symposium, eds. D.G. Crabb, et al., AIP Conf. Proc. 1149 (2009)
- [4] T. Wiss et al, NIM A556 (2006) 1
- [5] E. Steffens, W. Haeberli, Rep. Prog. Phys. 66, (2003) 1887
- [6] R.J. Philpott, NIM A259 (1987) 317
- [7] A.D. Roberts et al, NIM A322 (1992) 6
- [8] D.C. Dinge et al, NIM A357 (1995) 195
- [9] P.A. Schmelzbach et al, NIM 186 (1981) 655
- [10] T.B. Clegg et al, NIM A357 (1995) 212
- [11] W. Haeberli, NIM 62 (1968) 355
- [12] A.S. Belov, Ref.[2], p. 209
- [13] A. Zelenski et al, Ref.[2], p. 221
- [14] L.W. Anderson, NIM 167 (1979) 363
- [15] E.A. Hinds, NIM 189 (1981) 599
- [16] D. Tupa, Phys. Rev. A33 (1986) 1045
- [17] P.G. Sona, Energia Nucleaire 14 (1967) 295
- [18] A. Kponou et al, Ref.[2], p. 241
- [19] Polarized Sources and Targets, Proceedings of 9<sup>th</sup> International Workshop, eds. V.P. Derenchuk and B. Przewoski, World Scientific (2002)
- [20] T.B. Clegg, AIP Proc. 117 (1984) p.63
- [21] R. Engels et al, RSI 74 (2003) 4607
- [22] T.B. Clegg et al, NIM 62 (1968) 343
- [23] D. Krämers et al, NIM 220 (1984) 118
- [24] H. Jänsch et al, NIM A254 (1987) 7
- [25] K. Asahi, et al., Ref.[3], p. 90
- [26] T. Roser, Ref.[2], p. 15
- [27] D. P. Grosnick et al, NIM A290 (1990) 269
- [28] E. Steffens, Ref.[3], p. 80

### 7.1.4 H<sup>-</sup> Ion Sources

K.N. Leung, LBNL

**Ion Production Processes** Three types of ion sources can be identified, surface sources, volume sources, and hybrid sources in which the former are combined.

In surface production, ions are formed in a back scattering process [1] in which H atoms or ions strike the surface, dissociate into neutral H atoms, scatter back from the lattice and pick up an electron on the way out. H<sup>-</sup> can also be formed by a desorption or sputtering process [2] if there are H atoms adsorbed on the surface being struck. In both desorption and reflection processes the probability for H<sup>-</sup> ion formation increases if the work function of the surface is low, this being effected by covering with a partial monolayer of an alkali metal such as Cs, K, Na, Li. In most applications, the alkali metal is deposited by evaporation into the hydrogen discharge which supplies the atoms and ions being converted to H<sup>-</sup> at the target surface. The sheath potential at the surface sets the energy scale of the emerging ions.

The energy spectrum of the ions produced by a Mo target immersed in a hydrogen plasma is shown in Fig.1 [3].

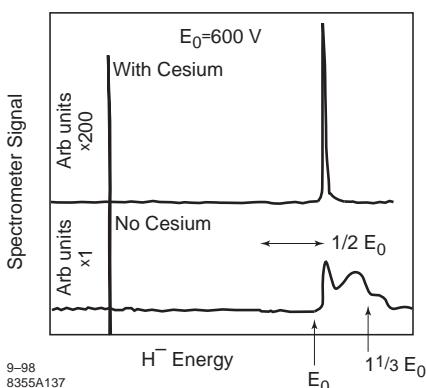


Figure 1: H<sup>-</sup> spectrum from Mo surface, with and without cesium.

The target is biased negatively with respect to the plasma. Since the plasma is generated by a low discharge power, the positive H species are dominated by H<sub>2</sub><sup>+</sup> and H<sub>3</sub><sup>+</sup>. For a pure hydrogen discharge the spectrum shows that there are four distinct groups of H<sup>-</sup> ions; the low energy desorption group and the three higher energy

reflection groups formed by the impact of H<sup>+</sup>, H<sub>2</sub><sup>+</sup>, and H<sub>3</sub><sup>+</sup> ions on the converter surface. For optimum Cs coverage, the energy spectrum of Fig.1 shows that desorption has become the dominant process and has been enhanced by more than 100 times. Mo seems to be the best target (converter) material. In volume production one has the processes of dissociative attachment [4], dissociative recombination [7, 8] and polar dissociation [8]. Dissociative attachment dominates in a typical hydrogen plasma. It is a two step process in which an ~100 eV electron excites H<sub>2</sub> into a vibrational state which then picks up an ~1 eV electron, thereby dissociating into an H<sup>-</sup> and H<sup>+</sup>, the cross section increasing rapidly with the vibrational state [5].

**Widely used ion source technology** Tab.1 shows some ion sources and their properties.

**New Ion Source Technology** The H<sup>-</sup> ions formed by a cesiated volume source have lower beam emittance and are therefore useful for the generation of high brightness beams. High current densities imply high discharge power, limiting filament life in dc or high duty service. This issue has been addressed in a new rf-driven multicusp source developed at LBNL [25]. The basic design is shown in Fig.2.

Today, small and large cesiated rf-driven H<sup>-</sup> sources are being employed in high energy accelerator systems (such as SNS) [26, 27] and MeV neutral beam injector for fusion research [28].

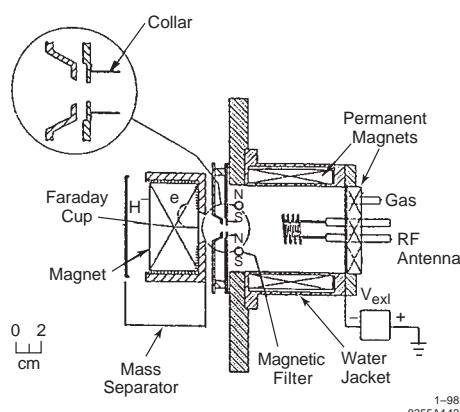
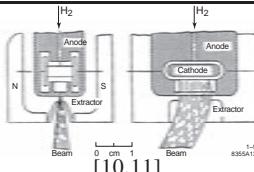
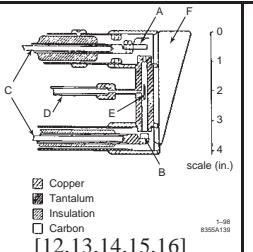
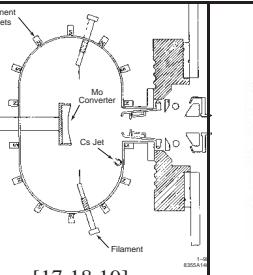
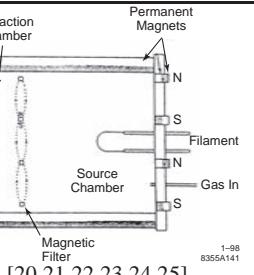


Figure 2: Rf driven multicusp volume source providing 35 kV, >35 mA at 0.1  $\pi$  mm-mrad in 100  $\mu$ s pulses at 10 Hz.

| Design                             |  |   |                         |   |
|------------------------------------|---|--|---|--|
| Name                               | Magnetron   | Penning  | Multicusp, surface  | Multicusp, volume  |
| I dc [mA]                          | -   | >5 (no Cs)   | -   | >10  |
| I pulse [mA]                       | 100   | 35, >100 (with Cs)   | 20  | 35 (no Cs), >120 (with Cs)   |
| $\tau$ [ms] pulse                  | ~100  | -  | 1   | 0.1-1  |
| Duty %                             | <1  | 1.2  | 12  | 0.1  |
| $\epsilon\beta\gamma \pi mm$ mrad* | 0.3-0.4   | 0.05   | 0.2   | 0.1 (no Cs)  |
| e / H-                             | 1   | -  | -   | >1 (no Cs), ~1 (with Cs)   |
| Discharge type                     | magnetron discharge   | Penning discharge  | dc discharge  | dc or rf induction discharge   |
| I discharge                        | 15  | 5  | 40  | >100 (pulsed), ~20 (dc)  |
| V discharge                        | 165   | 300  | 100   | 100  |
| gas flow sccm                      | 1   | 23   | -   | 10-20  |
| Cs mg/hr.                          | 0.5   | -  | -   | -  |
| Remarks                            | 1. hybrid source operation  | 1. operated as a volume production source in pure hydrogen discharge<br>2. operated as a hybrid source in the presence of cesium | 1. this is a pure surface-production source<br>2. I >1 A by employing a large converter and exit aperture | 1. operated as a hybrid production source in the presence of cesium<br>2. can be operated with rf-induction discharge in pulsed or CW mode<br>3. multi-ampere H- beams can be obtained by employing multiple extraction apertures. |

\* rms, normalized emittance

## References

- [1] J.R. Hiskes, A. Karo, M. Gardner, *J. Appl. Phys.* 47 (1976) 3888
- [2] M. Siedl, A.N. Paragellis, *Bull. Am. Phys. Soc.* 23 (1978) 804
- [3] K.N. Leung, K.W. Ehlers, *J. Vac. Sci. Tech. A3(3)* (1985) 1240
- [4] M. Allan, S.F. Wong, *PRL* 41 (1978) 1795
- [5] J. Wadehra, *PRA* 29 (1984) 106
- [6] B. Peart, K.T. Dolder, *J. Phys. B* 8 (1975) 1570
- [7] B. Peart, R.A. Forrest, K. Dolder, *J. Phys. B* 12 (1979) 3441
- [8] G.J. Shulz, R.K. Asundi, *PR* 158 (1967) 25
- [9] Yu I. Bel'chenko, G.I. Dimov, V.G. Dudnikov, *Investiya of USSR Academy of Science Ser. Fix.* 37 (1973) 2573
- [10] Proc. H<sup>-</sup> ion source for high energy accelerators workshop (1992) (unpublished)
- [11] K.W. Ehlers, B.F. Gavin, E.L. Hubbard, *NIM* 22 (1963) 87
- [12] V.G. Dudnikov, Proc. IV All-Union Conf. Part. Acc., Nauka 1 (1975) 323
- [13] P.W. Allison, *PAC* 77, p.1594
- [14] H.V. Smith Jr., J.D. Sherman, P. Allison, Proc. Linac Conf., CEAFA-Report-89-001 (1989) 164
- [15] R. Sidlow et al, EPAC 96
- [16] K. N. Leung, K. W. Ehlers, *RSI* 53 (1982) 803
- [17] R.R. Stevens Jr., R.L. York, J. McConnell, R. Kandarian, Linac Conf., GSI-84-11 (1984) 226
- [18] Y. Mori, A. Takagi, K. Ikegami, S. Fukumoto, *Proc. Int. Symp. on the Production and Neutralization of Negative Ions and Beams* (1986)
- [19] K.N. Leung, K.W. Ehlers, M. Bacal, *RSI* 54 (1983) 56
- [20] D.H. Yuan et al, *AIP Proc.* 158, p.346
- [21] R.L. York et al, *RSI* 55 (1984) 681
- [22] K.N. Leung et al, *RSI* 59 (1988) 453
- [23] K.N. Leung et al, *RSI* 60 (1989) 531
- [24] K.N. Leung, S.R. Walther, W.B. Kunkel, *PRL* 62 (1989) 764
- [25] K.N. Leung et al, *RSI* 62 (1991) 100
- [26] K. Saadatmand et al, *RSI* 66 (1995) 3438
- [27] M.P. Stockli et al, *RSI* 81 (2010) 02A729
- [28] E. Speth et al, *Nucl. Fusion* 46 (2006) S220

### 7.1.5 Antiproton Production

*G. Dugan, Cornell U.*

(see also Secs.1.6.1, 7.1.3,7.2.12)

**Transverse plane considerations** The basic production reaction  $p+N \rightarrow p+N^*+\bar{p}+p+X$  has a forward yield that rises in proportion to the incident p energy, at high energy. For a given proton energy the antiproton forward yield rises with

antiproton energy to a peak [1, 2, 3]. The rise with incident proton energy favors high incident energy but the production rate depends on the cycling rate of the proton source which may decrease with proton peak energy resulting in an optimum targeting energy. The antiproton energy distribution impacts the design of the matching, transport systems and the collection ring.

**Thin target yield:** Fig.1 illustrates the production from a thin target.

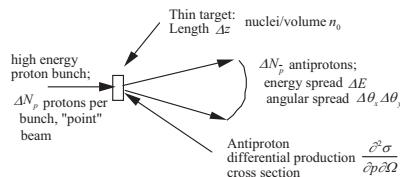


Figure 1: Thin target  $\bar{p}$  production.

The angular part of the emergent  $\bar{p}$  emittance is characterized by the rms production angle,  $\sigma_\theta = p_t/p$  which decreases with  $\bar{p}$  momentum since  $p_t$  is roughly constant at a few hundred MeV, typical of the momentum transfer in hadronic processes. The displacement part of the emittance is dominated by the incident proton beam size, putting a premium on small spots at the production target. For a Gaussian incident proton beam, the phase space density of the emergent  $\bar{p}$  is bi-Gaussian in  $\theta_{x,y}$  relative to  $\sigma_\theta$  and  $x, y$  relative to rms spot size at the target. Thus we may write for the forward  $\bar{p}$  phase space density

$$\frac{d^2}{dp dz} \bar{p}(0, 0, 0, 0) = \left[ \frac{dY}{dp} \frac{1}{\pi \sigma_\theta^2} \right] \frac{N_p}{2\pi \lambda \sigma_x \sigma_y} \quad (1)$$

where  $\lambda$  is the proton absorption length and  $Y$  is the yield. The numerical value for the Fermilab example shown in Tab.1 is  $\approx 300 \bar{p}/(\text{mm-mrad}^2/\text{GeV}/\text{cm})$ .

**Thick target yield:** With a target of finite thickness we need to account for the reabsorption of the  $\bar{p}$ . In terms of the target thickness  $L$  and the scaling factor  $\varsigma = L/\lambda$  and taking  $\lambda$  to be the same for p and  $\bar{p}$ ,

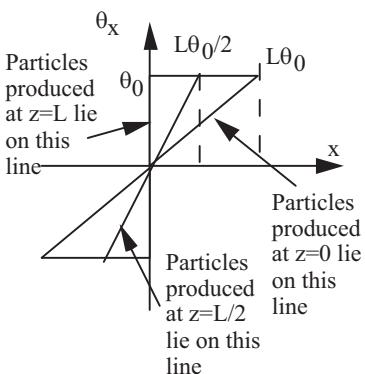
$$\frac{1}{N_p} \frac{dN_{\bar{p}}}{dp} (\varsigma) = \frac{dY}{dp} \varsigma \exp(-\varsigma) \quad (2)$$

which is maximum for  $\varsigma = 1$  and gives a numerical value of  $2.8 \times 10^{-4} \bar{p}/\text{proton}(\text{GeV}/c)$ . Note that when the full phase space density and finite target length are properly taken into account the

Table 1 FNAL  $\bar{p}$  source, original design [8]

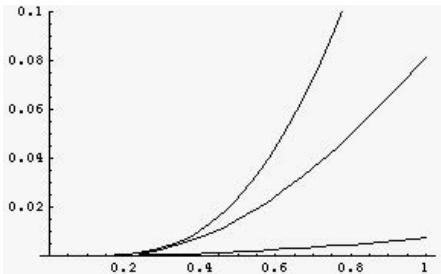
|   |   |
|---|---|
| $\lambda$   | 10 cm                                       |
| $\frac{dY}{dp} \frac{1}{\pi\sigma_\theta^2}$        | 0.118 $\bar{p}$ /steradian/<br>GeV/c/proton |
| $\sigma_\theta$                                     | 50 mrad.                                    |
| L   | 5 cm  |
| $N_p$   | $2.5 \cdot 10^{10}$                         |
| $\sigma_{x,y}$                                      | 0.04 cm                                     |
| $\sigma_t$ rms bunch length                         | 0.15 ns                                     |
| $\Delta E$ $\bar{p}$ energy spread<br>collected     | 0.04E=0.36 GeV                              |
| A <sub>x,y</sub> acceptance of<br>collection system | $20\pi$ mm mrad                             |
| $\bar{p}$ mean velocity                             | 0.994 c                                     |

Phase space at target end

Figure 2: Development of  $\bar{p}$  phase space in target.

optimum length is actually  $\varsigma \approx 1/2$ . For the Fermilab design, the accepted momentum bite is about 0.36 GeV/c and the collection efficiency is about 10%, giving a net  $\bar{p}/p$  ratio of  $10^{-5}$ .

Making the simplifying assumption that the angular distribution of the  $\bar{p}$  is uniform out to  $\sigma_\theta$  and is zero for larger angles, Fig.2 displays the resulting phase space distribution. The widening towards large  $x$  is called the “butterfly effect”. We see that in this approximation the effective source size will be about  $L\sigma_\theta/2$ . While useful for rough estimates, this approximation overstates the effective source size and masks the fact that the proton beam size plays a determining role down to proton beam sizes where multiple scattering in the target becomes important, about  $\sigma_r = 0.1$  mm in the parameter regime considered here. An exact solution for the  $\bar{p}$  phase space density from a

Figure 3:  $\bar{p}(r, s)$  vs.  $s$  for (lower to upper)  $r = 0.2, 0.5$ , and  $1.0$ .

thick target may be found from a transport equation [10]. The result is that the bulk of the  $\bar{p}$  lie within  $\pi\sigma_r\sigma_\theta$ . The central density is

$$\frac{d}{dp}\bar{p}(0, 0, 0, 0, \varsigma) = \frac{dY}{dp} \frac{N_p \varsigma \exp(-\varsigma)}{\pi^2 \sigma_\theta^2 \sigma_r^2} \quad (3)$$

This can be integrated over the admittance of the following transport line and collection ring to determine the number of  $\bar{p}$  per momentum that can be collected. For complex apertures, a Monte Carlo technique is often best.

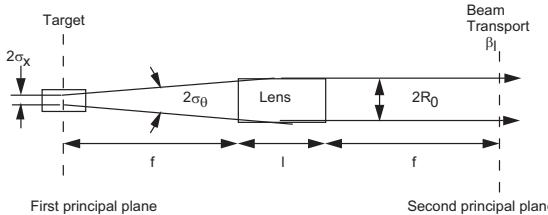
For numerical estimates of the yield, an analytical result [10] can be obtained with some simplifying assumptions. If both planes have the same maximum admittance,  $A_m$ , and the same  $\beta$ -function at the target,  $\beta$ ,

$$\frac{dN_{\bar{p}}}{dp} = N_p \frac{dY}{dp} \varsigma \exp(-\varsigma) \bar{p}(r, s) \\ s^2 = A_m / \sigma_r \sigma_\theta; \quad r = \sigma_r / \sigma_\theta \beta \quad (4)$$

Fig.3 presents the factor  $\bar{p}(r, s)$  for  $L\sigma_\theta/\sigma_r = 6.7$ . Small  $\beta$  (large  $r$ ) is favored.

In our example  $\sigma_r\sigma_\theta = 28$  mm-mrad,  $r = 1$ ,  $s = 0.85$  and  $\bar{p} = 0.12$ . The rate of accumulation in the collection ring is thus  $dN_{\bar{p}}/dp$  multiplied by the number of proton bunches per cycle,  $B = 80$  and the cycle rate  $\dot{C} = 0.5$  Hz. Carrying this out for the example here we obtain  $1.2 \times 10^{11} \bar{p}$  per hour per GeV/c.

Target Heating [4, 5]:  $dE/dx$  ionization losses in the target stress the target. It is found that if the energy deposited per gram exceeds 200-300 J then the target is subject to fracture. For this reason Cu is often used instead of W because of its relative ductility. This problem can be alleviated through use of a beam sweeping system [6] which scans the proton beam across the target during the pulse. The collection system must be swept in synchronism to avoid emittance dilution.

Figure 4:  $\bar{p}$  production target and matching lens.

**Longitudinal considerations [7]** The time duration of the  $\bar{p}$  pulse,  $\sigma_t$ , is essentially that of the  $p$  pulse so that the longitudinal density of the  $\bar{p}$  out of the target is  $\approx N_{\bar{p}}/\sigma_t \Delta E$ . To enhance this density, the proton bunch is rotated in phase space (Sec.4.10) just prior to extraction to achieve the shortest practical pulse. The rotation is accomplished by an rf voltage jump. The ratio of bunch lengths that can be achieved is roughly  $\sqrt[4]{V_{\min}/V_{\max}}$ ,  $V_{\max}$  being set by the rf system capability and  $V_{\min}$  by beam loading and instabilities. A factor of 3-4 in bunch length reduction has proved practical.

**Collection, transverse plane** The strongly divergent  $\bar{p}$  beam must be matched to the following transport line,  $\beta = \beta_l$  and ultimately to the collection ring. That can be done with a strong axisymmetric lens set for point to parallel focusing as shown in Fig.4.

For a lens of thickness  $\ell$  and gradient  $G$  the focal length for particles of momentum  $p$  is

$$f = 1/\sqrt{k} \tan \sqrt{k}\ell, \quad \text{where } k = eG/p \quad (5)$$

From Eq.(5) and Fig.4 we obtain the following relations which can be used iteratively together with the constraints of axisymmetric lens technology to design the lens and  $\bar{p}$  transport line:

$$R_0 \approx \frac{\sigma_\theta}{\sqrt{k} \sin \sqrt{k}\ell}; \quad \beta_l = \frac{x_l}{\theta_l} = \frac{2/Lk}{\sin^2(\sqrt{k}\ell)} \quad (6)$$

Axisymmetric lenses which have been used include lithium lenses (Sec. 7.2.12), horns [9] and plasma lenses (Sec.7.3.13). In all cases additional important considerations include minimizing  $\bar{p}$  absorption in the lens material, reduced multiple scattering and good lens quality. Because of its relatively good electrical conductivity and low  $Z$ , Li is used in the Fermilab case described above.

**Collection, longitudinal plane [7]** The collection lens, transport system and collection ring

have a combined energy full width  $\Delta E$  which defines the momentum spread accepted from the target. The accumulation ring downstream from the collection ring will generally have a smaller energy acceptance  $\Delta E_a$ . Hence it is necessary to reduce the momentum spread of the  $\bar{p}$  in the collector. This may be done by effecting a  $90^\circ$  rotation in longitudinal space followed by adiabatic debunching (Sec. 4.10). If the longitudinal emittance is preserved, then

$$\sigma_t \Delta E = (T_0/2B) \Delta E_{\text{fin}} \quad (7)$$

$T_0$  being the period of the collector and  $\Delta E_{\text{fin}}$  the energy spread of the debunched beam. If  $\Delta E_{\text{fin}} \leq \Delta E_a$  then the number of  $\bar{p}$  collected per cycle is

$$\approx B (dN_{\bar{p}}/dp) \Delta E \quad (8)$$

Note that  $B$  is not a free parameter. Fitting the bunches into the collector requires

$$B \leq h^* T_0 / T_0^* \quad (9)$$

where  $h^*$  is the harmonic number of the proton synchrotron and  $T_0^*$  its period. If  $E_a < \Delta E_{\text{fin}}$  then a further inefficiency is introduced. This may be avoided by adding some fast cooling to the collector to bring  $\Delta E_{\text{fin}} < \Delta E_a$ .

## References

- [1] C. Hojvat, A.J. van Ginneken, NIM 206 (1983) 67
- [2] I.L. Azhgirey et al, FNAL TM 1730 (1990)
- [3] S.C. O'Day, F.M. Bienosek, NIM A343 (1994) 343
- [4] Proc. High Intensity Targeting Workshop, Fermilab/U. Wisconsin (1980)
- [5] S. O'Day, F. Bienosek, K. Anderson, PAC 93, p.3096
- [6] F.M. Bienosek, K. Anderson, K. Fullett, PAC 95, p.1939
- [7] J. Griffin et al, PAC 83, p.2630
- [8] G. Dugan , XIII Int. Conf. on High Energy Acc. (1985) p.264
- [9] S. vander Meer, CERN PS/AA/80-12 (1980)
- [10] G. Dugan, CLNS 97/1532, Cornell U. (1997)

## 7.1.6 Multi-Charged Heavy Ion Sources

*J. Alessi, A. Pikin, BNL*

In heavy ion preinjectors, the choice of the ion charge state (charge-to-mass ratio) to be delivered from the ion source is an important design consideration. Higher Q/M from an ion source makes the downstream accelerators more compact and less costly, but generally there is a tradeoff between intensity and charge state from a source, which may or may not be acceptable. One might also select a charge state high enough to eliminate one or more subsequent stripping stages. The development of high charge state heavy ion sources is very active, since many existing and proposed heavy ion accelerators require sources with parameters pushing the state-of-the-art.

Source requirements depend strongly on the application. Important considerations are beam species (solid, gaseous), beam current, and duty factor required. Depending on the application, factors such as emittance, lifetime, reliability, stability (both pulse-to-pulse and long term), purity of the beam, gas load from the source, complexity, or ease and speed of changing species, can be important. One can refer to one of the excellent books on ion sources [1, 2], or published proceedings of ion source conferences, either general or for specific type sources.

For the production of high charge state ions, one needs high energy electrons in the source. A high density of these high energy electrons is required to produce high intensity and charge state. The ions must interact with the electron beam or plasma for a time long enough to reach the desired charge state through stepwise ionization. It is desirable to keep the background pressure as low as possible to minimize the recombination of ions.

For high current, high charge state applications, the charge state distribution of ions coming from the source can be an important consideration because the extraction and initial transport has to be designed to handle space charge from the total extracted current, which can be anywhere from a few times to a hundred times larger than the current in the desired charge state, depending on the ion source and charge state.

The most common sources for the production of multiply charged ions are discussed in the following sections. A few examples of performance, and general features of these sources are presented in Table 1. Normalized, rms emittances

for these sources are typically in the range of 0.1–0.3  $\pi$  mm mrad.

**Metal Vapor Vacuum Arc (MEVVA)** The MEVVA ion source is based on ion generation in a dense vacuum arc which sustains on metal vapor produced on the cathode hot spot of this discharge [3, 4, 5]. The cathode is a solid rod of material that contains the desired beam species. A high voltage pulse is first applied between the cathode and an electrode to the side of the cathode, resulting in a spark. This triggers the main low voltage, high current arc, which occurs between the cathode and anode.

This type source is simple and compact, and can produce currents of 10's of milliamperes, but only from metals, generally has low duty factor, and low charge states.

**Penning (PIG)** The PIG ion source is based on ionization by electrons which oscillate along an axial magnetic field between two negative electrodes, with an annular positive electrode (anode) placed between them. One of the cathode electrodes can be hot for higher electron emission. There are versions of the PIG ion source with ion extraction from the plasma either perpendicular to the magnetic field or coaxially with the magnetic field through the opening in the one of cathodes. Metallic ions can be produced from volatile metal-containing compounds, vapors, or by sputtering the metal targets in the vicinity of the discharge, using a working gas such as Xe [5, 6, 7, 8].

This source is compact, and produces tens to hundreds of microamperes of medium charge state ions.

**Electron Beam Ion Source (EBIS)** In an EBIS, multicharged ions are produced by stepwise ionization by a dense electron beam, while held in an electrostatic trap [9]. The electron beam provides both ionization and radial space-charge confinement of ions, while axial confinement is achieved by creating potential barriers using cylindrical electrodes on both sides of the ionization region, completing the topologically closed potential trap. Ions are held in the trap until the desired charge state is at the peak of a charge state distribution, at which time one end barrier is dropped and all ions are extracted. The charge states distribution in EBIS with pulsed ion injection is narrow because the charge state-reducing channels, which are typical for plasma ion sources, have very small contributions in EBIS. One has control of the energy of the ionizing electrons and the

Table 1: Features of multiply charged positive ion sources.

| Type                                   | Working substances  | Ion type (ex.)                         | Current  | Duty factor      | Pros  | Cons  |
|--|---|--|--|------------------|---|---|
| MEVVA [3, 4, 5]                        | Solid conductors  | U <sup>4+</sup>                        | 10's–100's mA's<br>20 mA (0.6 ms, 0.25 Hz)                     | Pulsed           | Does not require buffer gas; very high currents; versions with multiple cathodes                            | Sometimes noisy; sometimes poor pulse stability; low charge states; typically low duty cycle; lifetime limited by cathode erosion and coating of insulators |
| PIG (Penning) [5, 6, 7, 8]             | Gases, volatile compounds, sputtering of solids           | Xe <sup>15+</sup><br>U <sup>10+</sup>  | 55 μA<br>350 μA  | Pulsed<br>Pulsed | Easy to produce metallic ions by sputtering; good pulsed lifetime, also operates dc                         | Requires buffer gas; noisy; fast erosion of cathode at high power operation   |
| EBIS [9, 10]<br>ESIS [11]              | Solids, gases (material often from external source as 1+) | Au <sup>32+</sup><br>Fe <sup>24+</sup> | mA's (~ 10 μs, 5 Hz)<br>150 μA (~ 8 μs)                        | Typically pulsed | High currents at high charge states; can produce any ion species; stable, can switch species pulse-to-pulse | Limited capacity of ion trap; complex; large; ultrahigh vacuum requirements   |
| ECR [12, 13]                           | Solids, gases   | Xe <sup>30+</sup><br>U <sup>35+</sup>  | 117 μA<br>175 μA   | dc<br>dc         | Operates dc or pulsed; stable; easy operation for lower charge states                                       | Highest charge states are low fraction of total beam; has “memory effect”; magnets and rf systems can be complex  |
| Laser ion source [14, 15]<br>DPIS [16] | Solids  | Pb <sup>27+</sup><br>Al <sup>9+</sup>  | Few mA for few μs (≤ 1 Hz)<br>10's of mA's few μs (RFQ output) | Pulsed           | High current for intermediate charge states   | Low stability; low duty factor; short pulses; complex laser; target erosion (DPIS more stable, simpler laser)   |

time of exposure of trapped ions to the ionizing electrons, making the EBIS unique as a source of very highly charged ions.

Beam from an electron gun propagates into a strong solenoidal magnetic field (ex. 5–6 T superconducting solenoid), leading to a compression of the electron beam to current density up to 600 A/cm<sup>2</sup> or higher depending on the compression method. After passing through the solenoid, the electron beam is decelerated and stopped on an electron collector at the other end. Inside the solenoid are a series of isolated drift tubes for trapping of the ions. The working ions can be injected into the trap using either outside ion sources (usually of singly charged ions) or from gas/vapor within the trap. Ion injection, ion confinement for the desired charge state distribution and ion extraction are done by changing the

potential distribution on the drift tubes. One can control the rate at which ions are released from the trap, and therefore the width of the extracted ion pulse can be optimized for single turn injection into a synchrotron. The total number of extracted charges is limited by the loss of radial confinement when the space charge of the electron beam is fully neutralized by the accumulated ions, so maximum ion output requires high electron currents (1–10 A), and long trapping regions (up to 1.5 m) [10]. A weak dependence of ion beam intensity on the confinement time makes the output intensity of the ion beam almost independent of the charge state, which is a unique feature of EBIS. Vacuum in the trap of 10<sup>-10</sup> Torr or better is required in the trap region to reach maximum charge states and intensity in the desired beam.

Rather than dumping the ionizing electron beam after a single pass through the trap region of an EBIS, there is a version of EBIS which utilizes an oscillating electron beam between cathode and electron reflector in the magnetic field [11]. In this case, electron currents of  $\sim 10$  mA have been used to produce ion output equal to that achieved with  $\sim 0.5$  A in a conventional EBIS.

This is the only source, which does not have plasma and therefore has a separate independent control of the electron and ionic components. A powerful version of this type source is relatively large and complex, but produces the highest charge states of any type source, and ions have a narrow charge state distribution. Milliampere currents of ions can be produced in  $\sim 10$  microsecond pulses. Output is independent of ion species (any species injected into the EBIS as  $1+$  ions), and species can be switched on a pulse-to-pulse basis.

**Electron Cyclotron Resonance Source (ECR)**  
 The ECR is the most common type multicharged heavy ion source [12]. The ECR plasma chamber is placed in an axial magnetic mirror field configuration, produced by two solenoid coils, and a radial cusp magnetic field produced by a sextupole magnet. The superposition of these fields produces a minimum-B configuration in the center. Gas for the species of interest, or alternatively a buffer gas, is fed into the chamber. Plasma is produced via the injection of microwave power into the chamber, and is confined in a magnetic field with a bottle configuration. Electrons in this region are resonantly heated to high temperatures, necessary for the high charge state ion production in the plasma, which then occurs predominantly via stepwise ionization. Ion current can be increased with increasing rf power, or by increasing both the rf frequency and magnetic field (current  $\sim$  scales with  $f^2$ ). When the beam of interest cannot be produced from a gas, the options are the heating/sputtering of solid material inserted into the plasma, using volatile metallic compounds, or the use of a very high temperature oven.

In pulsed operation, the ions can be extracted right after the rf power is turned off, called the “afterglow” mode. In this mode the peak ion current can be a factor of  $\sim 2$  higher than in the continuous mode, explained by the increase in the rate of electron loss when rf power is turned off, causing ions, confined by the electron space charge, to exit the plasma more quickly as well.

ECRs continue to move to higher frequencies and higher magnetic fields, in order to get higher currents and higher charge states. State-of-the-art ECR sources [13], now operate at frequencies up to 28 GHz, rf powers of 10–15 kW, and sometimes multiple frequencies are used, requiring multiple rf sources. The required superconducting solenoid and sextupoles push the state-of-the-art in superconducting magnet technology (very high Lorentz forces between solenoid and 6-pole, extreme tensions in coils supports).

An ECR is essentially the only choice when dc operation is required for high current, high charge state beams. The broad charge state distribution and need for a buffer gas means that the beam of interest might be only 1% of the total extracted current. For lower charge states and intensities, ECR’s get simpler (lower frequency, lower magnetic field, room temperature magnets).

**Laser Ion Source (LIS)** A short pulsed, high power laser beam is focused to a small spot on a solid target containing the desired beam species. Evaporation of target material occurs, and electrons in the plasma absorb laser energy via inverse Bremsstrahlung, causing their energy to increase. Plasma forms and rapidly expands normal to the target. Plasma ions are stepwise ionized to high charge states. The electron temperature increases with laser power density and wavelength, which is advantageous for the production of high charge states. This type source can produce very high currents in very short pulses [14, 15]. Because the peak current is often too high to transport due to space charge blowup, the plasma is usually allowed to expand for several meters before extraction of ions to reduce the peak current and increase the pulse width to some microseconds. Difficulties can come from target erosion, pulse-to-pulse beam current fluctuations, and coating of optics by target material. The laser required for this type source is large and expensive.

A variation which produces very high currents of highly charged, low to intermediate mass ions, using a much simpler laser is the direct plasma injection scheme (DPIS) [16]. In this scheme, a compact Nd:YAG laser can be used to ablate and ionize a solid target, the ions are allowed to drift as a plasma right to the RFQ entrance (eliminating problems coming from space charge during extraction and transport).

## References

- [1] The physics and technology of ion sources, 2nd, Revised and Extended edition, Ian G. Brown (Editor), Wiley, 2004
- [2] Handbook of Ion Sources, Edited by Bernard Wolf, CRC Press, 1995
- [3] I.G. Brown et al, Applied Physics Letters 47 (1985) 358
- [4] R. Hollinger et al, RSI 75 (2004) 1595
- [5] R. Hollinger et al, RSI 79 (2008) 02C703
- [6] T. Miyata et al, RSI 75 (2004) 1863
- [7] V.B. Kutner, RSI 65 (1994) 1039
- [8] P. Spädtke et al, NIM B139 (1998) 145
- [9] E.D. Donets, Physica Scripta T3 (1983) 11
- [10] J.G. Alessi et al, RSI 81 (2010) 02A509
- [11] D.E. Donets et al, JINST 5 (2010) C09001
- [12] R. Geller, IEEE Trans. Nucl. Sci. 23(2) (1976) 904
- [13] C. Lyneis et al, RSI 81 (2010) 02A201
- [14] A. Balabaev et al, RSI 75 (2004) 1572
- [15] B.Yu. Sharkov et al, EPAC08 (2008) p354
- [16] M. Okamura et al, RSI 79 (2008) 02B314

### 7.1.7 Foil Strippers

#### 7.1.7.1 Charge state strippers

*M.A. McMahan, LBNL*

For heavy ion accelerators, in order to achieve higher energies, or to run at lower power in rf systems, either a high-charge state source (ECR, EBIS) is used, or stripping is achieved through use of gas, solid, or other stripping mediums.

**Charge-changing cross sections** When an ion of charge  $q$  passes through matter, the ion may capture or lose electrons. For such a process, the probability for changing to a charge state  $q'$  is described by a cross section  $\sigma(q, q')$ . In thin gas targets (for which the density is low enough such that a collision-excited ion has time to return to its ground state before another collision) a system of coupled differential equations can be written [2],

$$\frac{\partial Y_q(x)}{\partial x} = \sum_{q' \neq q} [\sigma(q' - q)Y_q(x) - \sigma(q - q')Y_{q'}(x)] \quad (1)$$

where  $Y$  denotes the fraction of ions of charge  $q$ ,  $(\sum_q Y(q) = 1)$  and  $x$  is the target thickness.

These equations can be solved numerically or, for systems with only a few charge states, analytically, to give the non-equilibrium charge state distribution,  $Y_q$ . The cross sections for electron

capture,  $\sigma_c$  and electron loss,  $\sigma_l$  have been calculated theoretically by several authors [2]. Electron capture dominates at low energy/nucleon ( $E/A < 0.5$  MeV/u) when the velocity of the ion is comparable to the orbital velocity of the electrons. At higher energies ( $E/A > 2$  MeV/u), electron loss becomes the dominant feature.

**Equilibrium charge state distributions** An equilibrium thickness,  $t_{eq}$ , is reached when the cross section out of a charge state is equal to the cross section into it. In this case the distribution is

$$F(q) \approx \left(1/\sqrt{2\pi}d\right) \exp \left[ -(q - \bar{q})^2 / 2d^2 \right] \quad (2)$$

with the average charge state

$$\bar{q} = \sum_q qF(q) \quad \text{and} \quad d = \left[ \sum_q (q - \bar{q})^2 F(q) \right]^{1/2}$$

The degree of asymmetry of the distribution can be approximated by the skewness parameter  $s = \sum_q (q - \bar{q})^3 F(q) / d^3$ .

The use of solid foils for stripping give significantly higher  $\bar{q}$  than gases. This “density” effect occurs when the collision rate is greater than the rate for de-excitation of the ion from excited states [3, 4]. Since  $\sigma_l$  is greater for excited ions, this leads to higher charge states.  $\bar{q}$  begins to decrease for solid foils as one goes significantly above  $t_{eq}$  because of energy loss effects. For the highest charge states with minimal energy straggling, foil thickness should be chosen to be just above  $t_{eq}$ .

The type of stripper to be employed depends very much on the beam energy and intensity as well as the necessary charge state. Properties of the most common types of gas and solid strippers are given in Tab.1.

**Other strippers** Other strippers have been made employing jets of heavy particles at their vapor pressure, such as Hg or Fomblin [5] Y-Vac diffusion pump oil [6]. Fomblin, a perfluoropolyether with average molecular weight of 3400 give higher average charge states than standard gas strippers. In addition, charge state equilibrium occurs at vapor pressures of two orders of magnitude less than the pressure needed in standard gas strippers [7]. Plasmas have also been explored as strippers [8, 9]. Ionized plasmas have been shown to deliver charge states equal or greater to solid foils with less degradation of the beam emittance due to multiple scattering.

Table 1. Properties of common strippers

|   | Gas  | Foil   |
|---|--|--|
| + side  | 1) minimize thickness<br>2) less beam degradation<br>3) can take higher beam currents  | 1) can be made thicker for higher energy beams<br>2) higher charge states achieved (density effect)  |
| - side  | 1) must be windowless (need good vacuum)<br>2) limited in thickness  | 1) lifetime at high intensities<br>2) heat dissipation<br>3) larger effect on the beam emittance than gas  |
| Matl's  | He, N <sub>2</sub> , air, Ar, Kr   | Be, C, Al, Ta, Au (C most common)  |
| $t_{eq}$<br>( $\mu\text{gm}/\text{cm}^2$ )<br>[23] <sup>a,b,c</sup> | $t_{eq} \approx 1.66 A_p 10^{((0.076+0.11\log(E/A_p))Z_p + 0.803\log(E/A_p)-1.22)}$<br>[23] <sup>a,b,c</sup>   | $t_{eq} \approx 1.66 A_p 10^{((0.142+0.00831\log(E/A_p))Z_p + 0.824\log(E/A_p)-0.92)}$<br>[23] <sup>b,c</sup>  |
| $\bar{q} / Z_p$   | $\begin{aligned} &= \frac{376x + x^6}{1428 - 1206x^{0.5} + 690x + x^6} \\ &\text{for } x = (v/v_0 Z_p^{-0.52} Z_t^{0.03-0.017Z_p^{-0.52} v/v_0})^{1+0.4/Z_p} \quad [11] \end{aligned}$ | $\begin{aligned} &= (1 - \exp^{(-1.076X + 0.0836X^2 - 0.0335X^3)}) * \\ &(1 - 0.0019(Z_t - 6)X^{0.5} + 10^{-5}(Z_t - 6)^2 X) \text{ for} \\ &X = 83.3Z^{-0.45} \beta \quad [24]^d \end{aligned}$ |
| d   | $d = 0.35 Z_p^{0.55} \left\{ \bar{q} / Z_p [1 - \bar{q} / Z_p] \right\}^{0.27} \quad [25]$   | $d = 1.41 Z_p^{0.41} (\bar{q} / Z_p)^{0.3} [1 - (\bar{q} / Z_p)]^{0.37} \quad [26]$  |

$$\beta = v/c = \sqrt{1 - \left( \frac{1}{1+E/(A_p * 931.5)} \right)^2}; v_0 = 3.6 \times 10^6 \text{ m/s}; c = 3 \times 10^8 \text{ m/s}$$

<sup>a</sup> For gas cells,  $t = 9.66 * 10^{18} L P / T$  (molecules/cm<sup>2</sup>) =  $16.3 M L P / T$  ( $\mu\text{gm}/\text{cm}^2$ ), where  $L$  = cell length (cm),  $P$  = pressure (Torr),  $T$  = temperature (K), and  $M$  = molecular weight.

<sup>b</sup> least squares fit to experimental data

<sup>c</sup>  $t_{eq}$  exhibits a small dependence on the initial charge state of the ion which is not included in this approximation

<sup>d</sup> Parameters updated to fit data through 1992 by McMahan et al. [9]

**Shell effects** The semi-empirical formulae for  $\bar{q}$  given in Tab.1 seem to be the most useful of many that have been published over the years. These formulae were chosen because they cover the widest range of data, but the smooth trends they predict as a function of the ion energy  $E$  and atomic number  $Z_p$  are only averages which do not include the effects of electronic shells. The shell effects can have a large effect on both the average charge state and width, particularly for very heavy ions, where the difference in  $\bar{q}$  can be as large as five charge states. Attempts have been made to calculate correction factors for shell effects; however, these corrections are very dependent on both the ion and energy, and extrapolation is difficult [10, 11, 12]. It is best to resort to measurements in the region of interest. Compilations of data are available in [11] (before 1973), [13] (1973-1986), and [14] (1986-1992).

**Highly charged ions** In the high velocity region, where  $q/Z$  is approaching unity, the charge state distribution is no longer Gaussian. For cases in which  $F(Z_p) + F(Z_p - 1) \geq 0.97$ , the following empirical expression holds [15]:

$$F(Z_p - 1) / F(Z_p) = \exp(A' - X') \quad (3)$$

where

$$X' = 143 Z_p^{-0.48} \beta (1 - \beta^2)^{1/2}$$

$$A' = 0.67 (Z_p - 3)^{0.18} + 0.3 \text{ with } \beta = v/c$$

$$= \sqrt{1 - 1/(1 + E/931.5A_p)^2}, E \text{ in MeV}$$

**Lifetimes of carbon foils** The lifetime of a C foil stripper is an important consideration. Foil lifetimes under ion bombardment appear to be limited by structural changes induced by radiation damage and possibly beam heating. Foil lifetimes can be estimated by [16]

$$T_{\text{foil}} = k_{\text{foil}} E / Z_p^2 M_p, \quad [p \mu\text{A min mm}^{-2}] \quad (4)$$

with  $E$  in eV and  $p$  the number of particles.  $k_{\text{foil}}$  depends on the detailed microstructure of the foil and depends on manufacture technique. Typical values are  $k_{\text{foil}} = 0.018$  for foils produced by vapor deposition and  $k_{\text{foil}} = 0.0073 t - 0.010$  for foils of thickness  $t$  [ $\mu\text{g}/\text{cm}^2$ ] made by glow discharge cracking of ethylene with 10% Ar and 2.5 kV bias.

Lifetimes of C foils can be improved by various techniques, including the mounting of slack foils, laser annealing, and pre-electron bombardment [17].

Table 2. Some Examples of Strippers used in Heavy Ion Accelerators

| Accelerator                                   | Energy<br>(MeV/u) | Stripper   | q of typical<br>beam                 | Ref. |
|---|-------------------|--|--------------------------------------|------|
| ETH/SIN AMS Facility<br>[tandem]              | 0.1-0.4           | $\approx 3 \mu\text{gm}/\text{cm}^2 \text{ O}_2 \text{ or Ar}$ | C:<br>$-1 \rightarrow 4$             | [27] |
| AGS (BNL)[synchrotron]<br>Tandem terminal:    | .07-.2            | $2-5 \mu\text{g}/\text{cm}^2 \text{ C foil}$                   | Au:<br>$-1 \rightarrow 12$           | [28] |
| Tandem exit:                                  | 0.93              | $12 \mu\text{g}/\text{cm}^2 \text{ C foil}$                    | $12 \rightarrow 32$                  |      |
| Booster exit:                                 | 95                | $4-10 \text{ mil C foil}$                                      | $32 \rightarrow 77$                  |      |
| RIA (proposed)<br>After low- $\beta$ section: | 9.3               | Li film or C foil  | $U:^\#$<br>$28-29 \rightarrow 69-73$ | [29] |
| After medium- $\beta$ section:                | 80.3              | C foil   | $69-73 \rightarrow 87-90$            |      |

<sup>#</sup> multiple charge states to be accelerated

Table 3 The functions  $g_1(\tau)$  and  $g_2(\tau)$  [14]

| $\tau$ | 0.2   | 0.4   | 0.6   | 0.8   | 1.0   | 1.2   | 1.4   | 1.6   | 1.8   |
|--------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| $g_1$  | 0.05  | 0.115 | 0.183 | 0.245 | 0.305 | 0.363 | 0.419 | 0.473 | 0.525 |
| $g_2$  | -     | 1.25  | 0.91  | 0.79  | 0.73  | 0.69  | 0.65  | 0.63  | 0.61  |
| $\tau$ | 2.0   | 2.5   | 3.0   | 3.5   | 4.0   | 4.5   | 5.0   | 6.0   | 7.0   |
| $g_1$  | 0.575 | 0.689 | 0.799 | 0.905 | 1.01  | 1.10  | 1.19  | 1.37  | 1.54  |
| $g_2$  | 0.59  | 0.56  | 0.53  | 0.50  | 0.47  | 0.45  | 0.43  | 0.40  | 0.37  |
| $\tau$ | 8.0   | 9.0   | 10    | 12    | 14    | 16    | 18    | 20    |       |
| $g_1$  | 1.70  | 1.85  | 1.99  | 2.27  | 2.54  | 2.80  | 3.05  | 3.29  |       |
| $g_2$  | 0.34  | 0.32  | 0.30  | 0.26  | 0.22  | 0.18  | 0.15  | 0.13  |       |

**Effects on beam** The use of a stripper can have a negative effect on the beam emittance through multiple scattering and energy straggling. This effect is largest for solid foils and/or low energy beams. Meyer [18] calculates the angular spread of the beam in terms of an energy-independent reduced half width which is valid in the region  $E/A < 0.0248 Z_P^2 Z_T^2$ ,

$$\bar{\theta}_{1/2} = g_1(\tau) + (a^2/r_0^2) g_2(\tau) \quad (5)$$

where the screening parameter

$a = 0.885 a_\infty / \sqrt{Z_P^{2/3} + Z_T^{2/3}}$ , with the Bohr radius  $a_\infty = 0.529 \times 10^{-8} \text{ cm}$ . The reduced thickness,  $\tau = \pi a^2 t / 1.6605 \times 10^{-18} A_T$  with  $t$  in  $[\mu\text{gm}/\text{cm}^2]$  and  $r_0 \approx 0.5 N^{-1/3}$ ,  $N$  being the atomic density. The second term in Eq.(5) is generally a small correction to  $g_1(t)$ . Values of  $g_{1,2}(t)$  are in Tab.3. The Meyer formulation has been validated by many experiments. For higher energies scattering effects are of less importance, but if estimates are needed, the theory of Moliere [19] can be used.

For  $E/A \approx 1 \text{ MeV}$ , an estimate for the energy straggling due to the stripper can be made as

(after [20] with parameters refined in [21]):

$$\delta E = 0.0425 Z_P \sqrt{\frac{Z_T}{A_T}} t \text{ [MeV]} \quad (6)$$

with  $t$  in  $[\mu\text{gm}/\text{cm}^2]$ . For specific target/projectile and energy combinations, modern ion stopping Monte Carlo codes such as SRIM 2003 [22, 23] can be used to give an accurate calculation of both the angular and energy straggling.

The field of high intensity, heavy ion beams is advancing rapidly and with it the stripping technology needed to go with it. A survey paper describing the alternatives under study now can be found in auxiliary reference [1].

## References

- [1] F. Marti et al, LINAC 10, TUP106, BNL (2010)
- [2] H.D. Betz, Rev. Mod. Phys. 44 (1972) 465
- [3] N. Bohr, J. Linhard, Kgl. Danske Videnskab. Selskab, Mat-Fys. Medd. 28 (1954) 7
- [4] G. Ryding, H.D. Betz, A.B. Wittkower, PRL 24 (1970) 123
- [5] marketed by Montedison USA
- [6] J.R. Alonso, B.T. Leeman, PAC 79, p.3718

## Sec.7.1: PARTICLE SOURCES

- [7] R.B. Clark, I.S. Grant, R. King, D.A. Eastham, T. Joy, NIM 133 (1976) 17
- [8] G.D. Alton, R.A. Sparrow, R.E. Olson, PRA 45 (1992) 5957
- [9] Y. Oguri et al, NIM B161 (2000) 155
- [10] M.A. McMahan, B. Feinberg, LBL-29758
- [11] Wittkower, Betz, Atomic Data 5 (1973) 113
- [12] G. Schiwietz, P.L. Grande, NIM B175 (2001) 125
- [13] K. Shima, T. Mikumo, H. Tawara, Atomic Data and Nucl. Data Tables 34 (1986) 357
- [14] K. Shima et al, Atomic Data and Nucl. Data Tables 51 (1992) 173
- [15] K. Shima, N. Kuno, M. Yamanouchi, PRA 40 (1989) 3557
- [16] R.L. Auble, D.M. Galbraith, NIM 200 (1982) 13
- [17] D.W.L. Tolfree, NIM 200 (1982) 15
- [18] L. Meyer, Phys. Stat. Sol. 44b (1971) 253
- [19] G. Moliere, Z. Naturforsch 2a (1947) 133; Ibid, Z. Naturforsch 3a (1948) 78
- [20] N. Bohr, Kgl. Dan. Vid. Selsk. Mat.-Fys. Medd. 418 (1948) 8
- [21] S. Ouichaoui et al, NIM B164 (2000) 259
- [22] J.F. Ziegler, J. Appl. Phys./Rev. Appl. Phys. 85(1999) 1249
- [23] www.srim.org (2003)
- [24] V.P. Zaikov et al, NIM B5 (1984) 10
- [25] K. Shima, T. Ishihara, T.M. Kumo, NIM 200 (1982) 605
- [26] R.O. Sayer, Rev. de Phys. Appl. 12 (1977) 1543
- [27] H.D. Betz, in Applied Atomic Collision Physics, Vol.4, Academic Press (1983) p.1
- [28] H.J. Hofmann et al NIM B29 (1987) 100
- [29] P. Thieberger, BNL, private communication
- [30] RIA Facility Pre-conceptual Design report; www.orau.org/ria/pdf/p-1-0-4.pdf (2003)

### 7.1.7.2 Stripper foils for H<sup>-</sup> beams

*M.A. Plum, ORNL*

Stripper foils for high intensity H<sup>-</sup> beams present special challenges:

- 1) Foil temperatures can be high enough to evaporate (sublimate) and/or melt the material.
- 2) Partially stripped H<sup>0</sup> beams can be in excited states (H<sup>0\*</sup>), where the electron is even more loosely bound to the nucleus and thus susceptible to stripping, due to magnetic fields, in regions that are outside the acceptance of the ring, thus leading to beam loss.
- 3) Charge developed on the foil due to, e.g., secondary electron emission and thermionic

electron emission. If the charge exceeds a certain threshold, the result could be vacuum breakdown (cathode-spot in-vacuum breakdown and anode-spot in-vacuum breakdown), which can damage the foil and the bracket [1]. This is especially important for foils that have high resistivity.

- 4) The beam power in the electrons stripped from the incoming H<sup>-</sup> beam can be high enough to require special handling (e.g the electrons stripped from a 1 MW H<sup>-</sup> beam have an electron beam power of 1.1 kW).

Most high intensity charge-exchange-injection machines rely on carbon-based stripper foils due to their low vapor pressure; high melting temperature; and their low atomic number ( $z = 6$ ), which allows a relatively small contribution to beam scattering that can cause beam emittance blow up and beam loss. Carbon foils are easily fabricated, handled, and mounted; and they are also well suited to vacuum applications. Other foil types, such as aluminum oxide, have also been successfully used [2].

**Stripping efficiency** The fractional yield of graphite stripper foils as a function of H<sup>-</sup> beam velocity has been parameterized to be [3, 4]

$$\begin{aligned} H^- &= \exp[-(a + b) \times d \times t] \\ H^0 &= \left( \frac{a}{a + b - c} \right) (\exp(-c \times d \times t) \\ &\quad - \exp[-(a + b) \times d \times t]) \\ H^+ &= 1 - H^- - H^0 \end{aligned}$$

where

$$a = 0.479 \times 10^{-18} \text{ cm}^2 / \beta^2$$

$$b = 0.0085 \times 10^{-18} \text{ cm}^2 / \beta^2$$

$$c = 0.187 \times 10^{-18} \text{ cm}^2 / \beta^2$$

$$d = \text{graphite density} = 5 \times 10^{16} \text{ atoms} / \mu \text{g}$$

$$t = \text{foil thickness in } \mu \text{ g/cm}^2$$

$$\beta = \text{relativistic factor at injection energy}$$

**H<sup>0</sup> excited states** H<sup>0</sup> excited states are created when an H<sup>-</sup> beam passes through a thin stripping foil with a stripping efficiency less than 100%. The H<sup>0</sup> excited states are populated according to  $n^{-2.8}$  [5], where  $n = 1, 2, 3, \dots$  is the principle quantum number of the H<sup>0</sup> atoms. The states with high values of  $n$  are easily stripped to H<sup>+</sup> by downstream magnetic fields [6] which are Lorentz-transformed into electric fields in the rest frame of the beam ( $E = \gamma\beta c B_{lab}$ ). Excited states with very short lifetimes (a few ps)

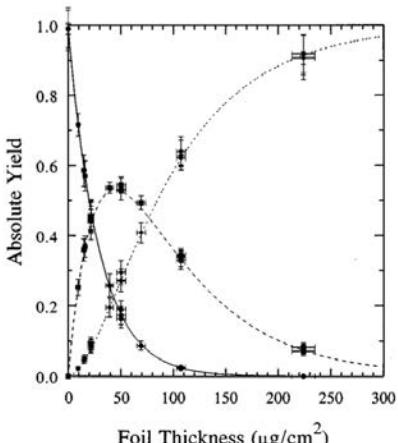


Figure 1: Stripping efficiency of carbon foil as a function of thickness for an 800 MeV  $H^-$  ion beam [5].  $H^+$  (dotted line),  $H^-$  (solid line), and  $H^0$  (dashed line).

will be stripped inside the acceptance of the ring and should not present a problem. Excited states with long ( $\sim 10$  ns) lifetimes will survive long enough to be properly delivered to a beam stop. Excited states with intermediate lifetimes can be stripped outside the acceptance of the ring but still be transported partly around the ring and cause undesirable beam loss. Excited state beam losses of up to 50% of the total beam loss have been observed [7]. The  $H^{0*}$  lifetimes and hence stripping probability can be computed using, for example, the parameterization in [8].

To manage the effects of  $H^{0*}$  states, it can be desirable [9] to place the stripper foil in a strong magnetic field to immediately strip the states with high quantum numbers (e.g.  $n \geq 6$ ) to ensure that these particles will be stripped inside the dynamic aperture. The downstream magnetic fields can then be set to values lower than the field at the foil to minimize the probability that any further stripping will take place prior to proper disposal of the remainder of the  $H^{0*}$  particles.

**Foil lifetime** Stripper foil lifetimes are notoriously difficult to predict. However, some authors have attempted to quantify the lifetime, e.g. [3, 10]. Relevant factors are primarily radiation damage, temperature, and thermal stress. An upper temperature limit is given by the vapor pressure, which is an exponential function of temperature:  $P(T) = Ae^{(-B/T)}$ , where  $A$  and  $B$  are constants. If the temperature is too high the foil

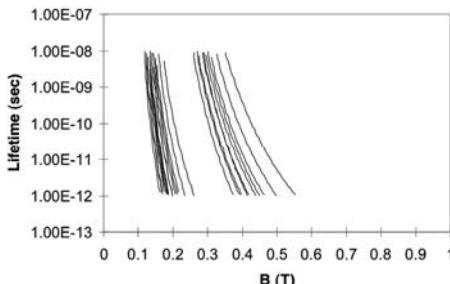


Figure 2: Lifetimes for 1000 MeV  $n = 4$  (band at right) and  $n = 5$  (band at left) excited state  $H$  versus magnetic field level.

simply evaporates. For example, for a carbon foil temperature of 2200 K, the evaporation rate is 1 micron/hr [10]. Evaporation rate calculations should take into account the time structure of the beam, since for pulsed beams the foil may be at high temperatures for only a small fraction of the time, which reduces the evaporation rate. Foil temperature calculations should include 1) temperature vs. time for the case of pulsed beams (temperature can vary more than 1000 deg. C between pulses); 2) delta-ray or knock-on electrons that lower the energy deposition in the foil; 3) emissivity, heat capacity, and heat conductivity as a function of temperature (note that these quantities are often not well understood at high temperatures); and 4) detailed treatment of the power deposited by the electrons stripped off the  $H^-$  beam and the neutral  $H^0$  particles.

### Foil fabrication technology

Many carbon-based foil fabrication techniques have been developed in the pursuit of long-lifetime foils, including evaporation, polycrystalline graphite, diamond-like carbon, AC-DC arc discharge, Hybrid Boron-Carbon (HBC), carbon-nanotube, and chemical-vapor-deposition (CVD) of microcrystalline and nanocrystalline diamond. The techniques that lead to the longest lifetimes at high temperatures include the AC-DC arc discharge [11], HBC [12] and nanocrystalline diamond [13]. New techniques are constantly being developed, such as boron-doping the CVD diamond foils and further refining the HBC technology.

**Other considerations** Other issues to be aware of include:

- 1) If the foil is placed in a magnetic field, the electrons stripped from the beam can circle

## Sec.7.1: PARTICLE SOURCES

around, pass through the foil multiple times, and add to the heating of the foil.

- If the foil is placed in a magnetic field, the electrons stripped from the beam can reflect from the electron beam dump and strike the foil, the foil bracket, and other surfaces [1].

## References

- M. Plum et al, IPAC10 (2010) p3971
- D. Findley, private communication, March 2010
- W. Chou, M. Kostin, Z. Tang NIM A 590 (2008) 1–12. W. Chou, private communication
- P. Kurpick et al, PR A58 (1998) 2183
- M. Gulley et al, PR A 53 (1996) 3201
- A. J. Jason et al, EPAC94 p1221
- D. Fitzgerald et al, PAC (1999) p518
- R.J. Damburg and V.V. Kolosov, J. Phys. B: Atom. Molec. Phys. 9(18) (1976) 3149
- J. Wei et al, PRST AB 3 (2000) 080101
- S.G. Lebedev, A.S. Lebedev, PRST AB 11 (2008) 020401
- I. Sugai et al, NIM A 362 (1995) 70–76
- I. Sugai et al, NIM A 561 (2006) 16–23
- R.W. Shaw et al, PAC07 p620

### 7.1.8 Lorentz Stripping of H<sup>-</sup> Ions

*M.A. Furman, LBNL*

When an H<sup>-</sup> ion moves in a magnetic field  $B$  it experiences a Lorentz force that bends its trajectory and also tends to break it up since the protons and electrons are bent in opposite directions, and the binding energy of the extra electron is only 0.755 eV. The breakup is a probabilistic process and quantum-mechanical in nature. In the ion rest frame, the stripping force is effected by the electric field  $E$  that is the Lorentz-transform of the magnetic field  $B$  in the lab,  $E = \kappa' \beta \gamma B$ , where  $\kappa' = 0.299792458$  GV/T-m. For the H<sup>-</sup> ion,  $E[\text{MV/cm}] = 3.197p[\text{GeV/c}] B[\text{T}]$  where  $p$  is the ion momentum in the lab.

The lifetime of the ion in an electric field can be calculated by applying the WKB approximation to the tunneling probability [2, 3, 4]. It has also been measured in several experiments [5, 6, 7, 8] whose results, for the ion's lifetime  $\tau$  in its own rest frame is well parametrized as

$$\tau = \frac{A}{E} \exp\left(\frac{C}{E}\right) \quad (1)$$

In the region of values of  $E$  where they overlap, the measurements in Refs.[5, 6, 7, 8] are fairly

consistent with each other, but are not consistent with [6]. Ref.[7], which covers the range  $E = 1.87 - 2.14$  MV/cm, has  $A = 7.96 \times 10^{-14}$  s-MV/cm and  $C = 42.56$  MV/cm, while Ref.[8], which covers  $E = 1.87 - 7.02$  MV/cm, has  $A = (2.47 \pm 0.09) \times 10^{-14}$  s-MV/cm and  $C = 44.94 \pm 0.10$  MV/cm. The mean stripping path length in the lab is given by

$$\lambda = c\beta\gamma\tau \quad (2)$$

Other mechanisms can strip H<sup>-</sup> ions, such as blackbody radiation, electromagnetic fields, interaction with residual gas [9], and intra-beam scattering [10, 11]. See Sec. 2.4.12.

## References

- M.A. Furman, D.E. Johnson, SSC-N-564 (1988), unpublished
- G. Darewych, S.M. Neamtan, NIM 21 (1963) 247
- B. Mullen, E.W. Vogt, Phys. Canada 24, No.3 (1968) 86
- L.D. Landau, E.M. Lifshitz, Quantum Mechanics (Nonrelativistic Theory), Pergamon (1958), Sec. 73
- S.N. Kaplan, G.A. Paulikas, R.V. Pyle, PR 131 (1963) 2574
- T.A. Cahill, J.R. Richardson, J.W. Verba, NIM 39 (1966) 278
- G.M. Stinson et al, NIM 74 (1969) 333
- A.J. Jason et al, LA-UR-81-1257 (1981)
- J.-P. Carneiro, R. Mustapha, P.N. Ostroumov, PRST AB 2, 040102 (2009)
- M. Chanal et al, PL B 192, 475 (1987)
- V. Lebedev et al, LINAC10, p931; PRL 108 114801 (2012)

### 7.1.9 Laser-Assisted H<sup>-</sup> Conversion to Protons

*V. Danilov, ORNL*

With the increase of average power of present high intensity proton rings and the rapid progress of laser technology, laser-assisted stripping has become a real alternative to carbon foils that are used for charge-exchange injection. After years of theoretical investigation of laser stripping feasibility (see, e.g., [1]), it was found that it is possible to overcome the main difficulty of the method – to excite hydrogen atoms with a very large spread of transition frequencies between the ground and some upper level of the hydrogen atomic beam using a narrowband laser [2]. High efficiency laser

stripping ( $\sim 90\%$ ), achieved experimentally at the Spallation Neutron Source in Oak Ridge, TN [3], paved the way to such devices.

**Main method** Hydrogen beam is obtained from an  $H^-$  beam after its transfer through a strong magnet. The energy spread ( $\sim 10^{-3}$ ), through the Doppler effect results in a large absorption line width as compared to the relative bandwidth of lasers ( $\sim 10^{-5} - 10^{-6}$ ). Conventional methods, such as Rabi oscillations, couldn't provide an excitation efficiency close to 100% for the typical linac beams.

The Doppler dependence of light frequency on incident angle and a convergent laser beam are utilized. By focusing the laser beam in the plane of the two beams, the angle of incidence of the laser light changes along the hydrogen beam path in the laser-particle beam overlap region. The laser frequency remains fixed, but because of the Doppler dependence of the rest-frame laser frequency on the incident angle, the frequency of the light in the atom's rest frame decreases as the angle increases. This introduces an effective frequency "sweep" as the hydrogen beam traverses the laser interaction region (it is conventionally called the adiabatic rapid passage).

The peak laser power was estimated in [2] for high efficiency stripping for the relativistic factor  $\beta \sim 1$ :

$$P_{peak} = \frac{\ln(1/\delta) \hbar^2 \varepsilon_0 c^2 \kappa \omega_0 h \sin \alpha}{2 \mu_{1n}^2 \gamma (1 + \beta \cos \alpha)^2}, \quad (1)$$

where  $\delta \ll 1$  is the ratio of unexcited to excited atoms,  $\alpha$  is the incident angle,  $h$  is the vertical FWHM of the laser beam,  $\omega_0$  is the laser frequency in the rest frame of the atom,  $\kappa$  is the full relative frequency change along the beam path, which, as follows from numerical simulations must satisfy  $\kappa \approx 6 \frac{\delta \gamma(rms)}{\gamma}$  of the beam to get high efficiency stripping,  $\mu_{1n} = \sqrt{\frac{2^{2n} n^7 (n-1)^{2n-5}}{3(n+1)^{2n+5}}} e a_\infty$ ,  $n$  is the upper level main quantum number, and the other letters ( $e, a_\infty, \dots$ ) correspond to the standard notations.

**Another method,** proposed by I. Yamane, to excite the levels with a large absorption line width includes widening the upper level with a magnetic field such that the level width is made to cover the transition frequency spread due to the Doppler effect (see [4] and references therein).

For the typical SNS linac parameters (assuming 90% of stripping),  $\beta \approx 0.875$ ,  $\alpha \approx 40^\circ$ ,  $\kappa \approx 3 \cdot 10^{-3} \omega_0$ ,  $h \approx 1 \text{ mm}$ ,  $n = 3$ , the formula (1) yields approximately 10 MW of peak laser power. There are several possibilities, such as dispersion function tailoring at the interaction point, a vertical size reduction, transition to mode locked lasers and the use of Fabri-Perot optical cavities [3] that can reduce the average laser power to a practical range below 100 W.

## References

- [1] I. Yamane, ICFA-HB2002, AIP Conference Proc. 642, Batavia, IL (2002) p.335.
- [2] V. Danilov et al, PRST AB 6, 5 (2003) 053501.
- [3] V. Danilov et al, PRST AB 10, 5 (2007) 053501.
- [4] T. Gorlov et al, PRSTAB 13, 5 (2010) 050101.

## 7.2 CONFINEMENT AND FOCUSING

### 7.2.1 Resistive Magnets

*F.E. Mills, D.J. Harding, FNAL*

(See also Secs.7.2.2, 7.2.3)

**Perfect iron, 2-D** Conceptual design begins by assuming  $\mu = \infty$  iron in a 2-D configuration in the  $x$ - $y$  plane and infinite extent in  $z$ .  $z$  is the beam direction, its axis at the center of the magnet aperture;  $x$  being the bend direction, positive outward; and  $y \perp z, x$ . In this case  $H = 0$  in the iron, and in the gap  $B = \mu_0 H$ . In the conductor-free gap  $B = -\nabla \psi$ , the gradient of a potential function, and the results of potential theory can be used (Sec.7.2.3.1). If the gap region is remote from the conductors, the iron surfaces are equipotentials, perpendicular to the field lines there. In general practice, the fields are described by a multipole expansion (Sec.6.17). Magnets may be designed to be pure multipoles or combined function magnets. Fig.1 shows the pole and yoke shapes of the most common iron magnets. Alternate poles are excited in alternate directions.

If  $a$  is the radius of the largest circle that can be drawn inside the poles, the ideal pole surfaces for multipoles and rotated (skew) multipoles of order  $n$ , expressed in polar coordinates, are

$$\rho^n |\sin n\theta| = a^n; \quad \rho^n |\cos n\theta| = a^n \quad (1)$$

$n = 1$  being the dipole, 2 the quadrupole, etc. For the simplest combined function magnet, a dipole

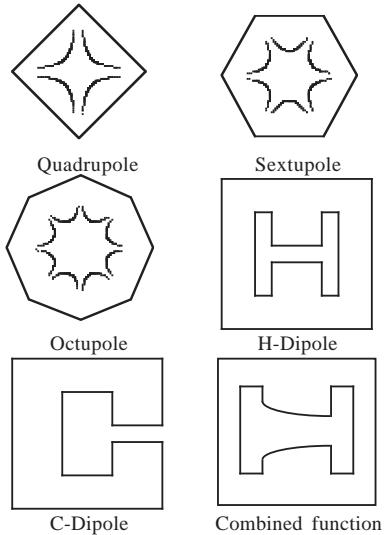


Figure 1: Multipole and combined function iron shapes.

with a gradient, the pole is given in Cartesian coordinates as

$$\left(\frac{\rho_0}{n} - x\right)y = \frac{\rho_0}{n}h_0 \quad (2)$$

where the origin of  $x$  and  $y$  is at the gap center,  $\rho_0$  is the radius of curvature of the equilibrium orbit,  $h_0$  is the half aperture at  $x = 0$ , and  $n$  is the field index, defined as  $n = g\rho_0/B_0$  for a central field  $B_0$  and a gradient  $g$ .  $\rho_0/n$  is sometimes called the gradient length. The pole shape can be further refined to introduce a sextupole component, as well, to control the chromaticity [1].

In the case of the dipole with half gap  $a$  and infinite  $\mu$  iron, the required current is given by Ampere's law to be  $I = B(2a)/\mu_0$ . For the quadrupole the equivalent relation is  $I = ga^2/2\mu_0$ .

**Finite permeability** (See also Sec.7.2.2)

We can estimate the effects of the finite permeability of real iron by applying Ampere's law along paths linking the current and passing through the air gap (Sec.7.2.2),

$$\oint \vec{H} \cdot d\vec{l} = \frac{B(2a)}{\mu_0} + H_{\text{iron}}l_{\text{iron}} = I \quad (3)$$

In an H-magnet, as one moves from center to edge of pole the path  $l_{\text{iron}}$  is reduced, increasing the  $B$ . This gives rise to a symmetric field error (sextupole, etc.), while in a C-magnet this leads to an asymmetric field error (quadrupole, etc.).

**Pole shaping** For economy one wants to limit the pole width. For magnets in which the field is much less than the saturation field ( $\ll 2$  T), one may simply terminate the pole in a  $90^\circ$  corner at  $x = \pm w/2$  where  $w$  is the pole width. By the boundary conditions, this requires the field to be infinite at the corner, but the finite  $\mu$  iron cannot comply perfectly. One can make an improvement by rounding the corner so that the field in the iron remains below saturation. Another strategy useful for high field magnets is to shape the pole so that the field in the iron is always in the  $y$  direction over the working region and is therefore uniform. Then the field shape in the gap will be the same at all fields and saturation effects will be minimized [2].

At the pole corner, the potential in the gap can not be uniform in  $y$ , even with perfect iron, since the equipotentials must bend  $90^\circ$  and thus the field acquires an  $x$  component. The errors in field so defined will propagate into the gap, thereby distorting the field, as discussed in Sec.7.2.3.2.

The width of the good field region in the gap can be enlarged by the addition of iron to narrow the gap at the edge region of the pole. This extra material is usually referred to as a shim or Rose shim [3, 4]. For precise calculation of the shim shape with the actual magnet and coil geometry and iron properties, one uses computer programs such as TRIM, POISSON, MIRT etc. [5] developed over the last 50 years by S.C. Snowdon, K. Halbach, W. Trowbridge and others. These observations apply to magnets of any multipolarity.

There is a comparable problem at the magnet ends. Here we are concerned that  $\int B dz$  be as independent of  $x$  as possible over the useful aperture. Were we to terminate the magnet in the  $z$  direction in a square end, then the integral field would decrease with increasing  $|x|$ , exhibiting multipoles, particularly the sextupole in a dipole magnet. By shaping the pole end, making the effective length longer as a function of  $x$ , one can correct this. Finding the shape can be done empirically by measuring and shaping iteratively. Today, 3-D computer programs such as OPERA [6] do an excellent job of modeling magnet ends. Of course the problem is exacerbated in high field magnets where flux emanating from the magnet end due to saturation also must be taken into account. For  $\mu = \infty$  iron the flux lines, normal at the exit from the iron, can approach the surface at grazing angle so there need be very little  $B_z$  in

the iron. As  $\mu$  decreases,  $B_z$  must increase. Even with some attempt to terminate the pole with, e.g., a compound bevel, these errors can reduce magnet strength by  $\sim 1\%$  for 1.7 T magnetic fields in the gap. If the magnets are all of the same length that is not a great issue. If several lengths are used in the lattice this effect must be taken into account. (Sec.7.2.2)

As mentioned, this problem can be alleviated if the magnet end profile is shaped so as to maintain the field in the iron in the  $y$  direction so no horizontal end flux is needed. For AC magnets this precaution has another benefit. Horizontal flux at the magnet end causes maximum eddy currents (Sec.7.2.3.3) and thus heating, which can be severe. This is minimized by minimizing the  $B_z$  at the end. For high  $\mu$  iron, one can derive a pole shape equation to accomplish this:

$$z = \frac{2a}{\pi} (1 + t); \quad y = \frac{2a}{\pi} \left( \frac{\pi}{2} + e^t \right) \quad (4)$$

where  $2a$  is the full magnet gap and  $t$  is a parameter (flux function) which varies between  $-\infty$  and  $+\infty$ . Such a shape can be approximated by use of several different laminations or one can bond blocks of laminations into “end packs”, the poles of which are subsequently machined to the desired contour with a numerically controlled milling machine.

The beam bends in a dipole magnet, of course, reducing the useful aperture if the magnet is straight. Rather than increasing the magnet size, the magnet can be designed to follow the curve of the particle orbit, minimizing the material cost at the expense of some increase in fabrication complexity and cost.

With curved magnets, several complications arise. First, since the laminations are stacked parallel, there is a reduction of available radial aperture as the bend angle gets large. This can be accommodated with a wider pole and coil window. Second, it is less expensive to fabricate the core with parallel ends. The core can then be stacked against a curved rail and the laminations periodically pressed to assure uniformity. The edge angle thus created can be easily compensated by the lattice quadrupoles. The effective edge angle, however, does not turn out to be half the bend angle but to be about 3 mrad less, per end, for each  $10^\circ$  of bend. This results from the fact that there is more flux on the outside than on the inside. This effect is dependent on the shims and end shapes of the poles.

## Core construction

### DC and slowly ramped

magnets The field quality of the magnet is determined by the pole shape, and to a lesser degree by the composition of the iron core. Low carbon steel, 1006-1008, in 1.5 mm (0.0598 in) thick laminations, generally provides a good balance of performance, cost, and ease of fabrication. The magnetic properties (permeability, saturation, and coercive force) depend on both the chemistry and heat treatment, so in cases where the performance is critical, the magnetic properties should be specified. If the magnet will be used in a high radiation area where activation is a concern, the steel vendor should be instructed to no use antimony, in spite of its beneficial effects on the grain size. The combination of a high neutron capture cross section and long lifetime of the resulting isotope make antimony a major source of residual activity even in minuscule quantities. [7] In a large production run, perhaps extending over months or even years, it must be emphasized to the vendor that consistency is more important than “improvement” of the magnetic performance.

The properties of two types of low carbon steel are shown in Fig.2 [7]. The Main Injector steel has a little bit of silicon and has been specially heat treated to grow its grain size, yielding a lower coercive force, visible in Fig.2, and a slightly lower permeability at high field. Since the Pbar Source rings operate DC, there is little need for low coercive force, and the cost was reduced substantially.

Punching sheet steel into laminations is generally cost effective when building multiple magnets, even in a DC magnet where the laminated structure may not be required. It is very difficult to obtain the same precision for this cost by machining forged steel pieces. DC magnets in

Two Steels at Low Field

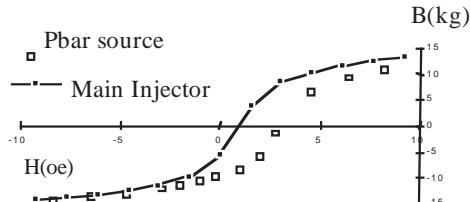


Figure 2:  $B$  vs  $H$  for two magnet steels.

small quantities, or those too large to employ laminations, are usually machined from solid steel blocks and assembled, saving the overhead of fabricating a stamping die.

Steel used for magnets is usually produced in wide sheets, typically 1-4 m in width, and rolled into a “master coil.” The steel is coated with a thin layer of electrical insulation. The choice of insulation material depends largely on the temperature that the core will be subjected to and, in some cases, the vacuum in which it will operate. The master coil is then slit into narrower coils, just slightly wider than the lamination size. The steel is not perfectly flat (thicker in the middle of the master coil than at the edges), with a crown of up to 1% of its nominal thickness in the center. Depending on the location across the coil, the lamination may be relatively flat, or have a taper. As such laminations are stacked, the magnet end surface becomes deformed unless one alternates the orientation of the taper to avoid the wrong taper at the end. A typical laminated H magnet core is made in two halves with mating surfaces. This requires a stacking fixture with a “pusher” to compress the laminations. The laminations are not perfectly flat because of the locked-in stress due to the punching. Further, the laminations are usually “flipped” every few inches to average out asymmetries due to die errors. Thus the laminations appear to be like springs. The stack is usually compressed until its effective Young’s Modulus exceeds  $1 \times 10^6$  psi [8] (The Young’s Modulus of steel is about  $3 \times 10^7$  psi.) Typical clamping pressure is 150-300 psi. The laminations are stacked on rails on their mating surfaces and pushed sideways against a rail. The laminations must be alternately pushed and compressed in order to position them properly.

Usually, end packs of solid steel or glued laminations are placed at each end to keep the stack from bowing out when the clamping pressure is released. Corrections for the end fields are included in these end packs.

The laminations and end packs are clamped to the table to produce a flat mating surface and minimize the field in the parting plane gap. This can be done with hydraulic systems, but a cheaper and more powerful method is to excite a magnetic field through the rails and the lamination back leg. Since the air gap in this loop can be small, a modest current can get a 1.5 T field, yielding a 88 psi stress between the lamination and the rail. Steel

bands, plates, or angle iron are welded to the laminations and end packs to hold the core together. It is critical that the design of the laminations and tie plates, along with the welding sequence that joins them, be engineered so that the “weld pull” works to produce a mating surface that is flat rather than wavy. A continuous weld bead pulling the laminations against a high precision reference surface is effective.

To assemble the magnet, the half cores are turned over, the coils are installed and fixed in place with any required ground insulation. The top half is rolled over, placed on the bottom half, the two are clamped together, and they are either bolted or welded together using tie plates. The welded magnet is superior, achieving much higher clamping stress than the bolted magnet, which usually is not clamped sufficiently to close the mating surfaces, adding reluctance to the magnetic circuit. If problems occur pulling the mating surface together, the back leg gap can be closed by magnetizing the core in the same flux pattern used on the stacking table.

C-magnet construction poses somewhat different problems. The laminations tend to be larger, so the problems of punching-induced stress and die errors are more serious. The laminations can only be flipped top for bottom, limiting the ability to average out errors. It is important to make the single back leg thick enough so that neither does the gap open up due to punching induced stress nor does the gap close due to magnet excitation. Attention must be paid to the crown effects as well. The magnets can be stacked with the gap straddling a vertical rail to position the laminations. They can be supported on curved rails to produce a curved magnet, if desired. Since there is no problem of mating surfaces, the laminations need only be clamped longitudinally during welding. Straps or plates should be welded on top and bottom to keep the laminations clamped during assembly and operation.

Quadrupole construction offers interesting choices. The magnet can be assembled from either two or four cores. It is much easier to maximize the conductor cross section in a four-core magnet, but the assembly of the cores into a whole magnet requires additional design effort and tooling to ensure the proper alignment of the poles with respect to each other and of the mating surfaces of the cores. Both approaches have been used successfully. Pole tip shims and shaped end

packs are used as in the dipole case. The use of a slotted weld strap on the outside allows post facto adjustment of the pole positions.

**Rapidly ramped and AC magnets** The biggest additional issues with AC magnet cores involve power losses from eddy currents and hysteresis. Eddy current losses in the core material call for thinner laminations and careful attention to avoiding shorts inside the core between laminations at their edges, where the steel coating can be bypassed. For example, a thicknesses of 0.35 mm (0.014 in) is appropriate for 60 Hz. In order to avoid the problems of the mating surface, it is preferable to build one-piece cores, either in H-magnet form, or in C-magnet form. Silicon steel, with silicon content 1-5%, is usually employed to reduce the hysteretic losses. The steel is non-grain-oriented [9]. A small penalty of  $\sim 0.1$  T is paid in saturation field by the use of silicon steel. Because of the tendency to operate AC magnets at lower fields ( $\sim 0.5$ - $0.9$  T) than DC or slowly ramped magnets, the saturation penalty is more than offset by the lower coercive force of silicon steel (0.5 Oe instead of 1-3 Oe for low carbon steel).

**Coil construction** The cost and reliability of the magnet is generally dominated by the coil and the power supply.

**DC and slowly pulsed magnets** Water cooled hollow copper conductor is usually employed to excite magnets with repetition rates  $\lesssim 30$  Hz. Normally the coil water configuration is chosen so that the water flow is turbulent to promote good heat transfer. Flow velocities should remain low to avoid erosion of copper. The water temperature rise is kept  $\lesssim 10^\circ\text{C}$ , with the main limitation being the internal stresses in the coil discussed below. At a flow of 1 gal/min and a temperature rise of  $3.8^\circ\text{C}$ , 1 kW of heat is removed by the water. For flow in a tube of length  $L$  (in) and diameter  $D$  (in) with a pressure drop of  $P$  (psi), the estimates for turbulent and laminar flows in gal/min are

$$F_{\text{turb}} = \sqrt{\frac{19,600 PD^5}{L}}, \quad F_{\text{lamin}} = \frac{31,900 PD^4}{L} \quad (5)$$

In any case, hole diameters less than about 3 mm (1/8 in) should be strictly avoided to reduce the chance of plugging with copper oxide and other debris. (See Sec.5.7) for a more detailed discussion of cooling by fluid flow.)

Three primary choices must be made regarding coil geometry: The total cross section of the coil, the size and shape of the individual

conductor, and whether to fill the mid-plane with conductor.

a) The cross section determines the copper cost and the total power dissipation. Analysis of construction and operating costs show that there is a broad minimum in overall cost for magnets with rms current densities in the region of 1.5-2.5 A/mm<sup>2</sup>.

b) The size of the conductor, whose multiple turns make up that total cross section, is chosen in concert with power supply considerations. More turns of smaller conductor require a lower current at a higher voltage due to both the resistive and inductive load. Smaller conductor is generally easier to wind into a coil, but the multiple turns are more labor-intensive to insulate and may not accommodate a water cooling passage of sufficient diameter.

c) Filling the whole space outside the gap with conductor reduces the steel cross section needed to accommodate the chosen conductor cross section. The disadvantage is that, rather than a simple "racetrack" coil, a "saddle" coil must be used, bending the conductor out of the aperture at the ends, which significantly complicates construction and uses more longitudinal space.

Where possible, coils should be wound from a single length of conductor to eliminate joints inside the magnet. Where joints are necessary, they must be designed, executed, and inspected with great care to avoid water leaks that can develop with time. A reliable technique is to counterbore each end of the conductor to be joined, then insert a ferrule with an inner diameter at least as large as the conductor water passage and braze the joint. The counterbore must be aligned carefully on the outside of the conductor, not to the hole (which does wander within the conductor), to ensure that the conductor ends are aligned.

A complication which occurs during coil winding comes when small radius of curvature bends are made in the coil [10]. The outside of the bend is stretched while the inside is compressed, so that the conductor acquires a trapezoidal shape. Winding under tension reduces this "keystoning", but does not remove it, and too much tension will tend to close the hole. Good practice is to bend with normal winding tension, and to limit the radius of curvature to no less than 2.5 conductor thicknesses. Under these conditions, the increase in conductor axial width and the reduction in

conductor radial width can be kept  $\lesssim 5\%$ . One must deal with even this amount. The keystoning will cause abrasion in the coil ends during insulating and curing, leading to turn-to-turn shorts later. The offending material can be ground away or the coil can be designed to include inter-turn spacers to separate the conductors by more than the keystoning. If space for the coil is not at a premium, the latter is usually less expensive.

For large, water-cooled coils the preferred insulation system is epoxy-impregnated cloth tape, usually fiberglass tape for its radiation resistance. After the coil is formed, the copper is wrapped with the glass tape (conductor wrap) and if there are many layers, the layers, or “pancakes” may be wrapped, and finally the coil receives a “ground wrap” to provide insulation from the iron core. The tape may be pre-coated with dry epoxy or may be clean during the wrapping. If the tape is pre-impregnated, then the coil is placed in a curing mold to set its shape and is cured in an oven. If plain tape is used, the whole system is vacuum impregnated with epoxy in either a mold or in the magnet core itself, then is cured. The coil insulation may be augmented with insulation attached to the core itself. The curing fixture is the main aid to precision in coil fabrication.

The performance of the epoxy insulation is key to the magnet reliability. The epoxy system is chosen for its dielectric properties, strength, flexibility, and resistance to radiation damage, making trade-offs that depend on the intended operating conditions. The insulation must withstand the water, and possibly other chemicals, to which it will inevitably be exposed. The conductor will expand and contract with its changing temperature, especially between periods of inactivity and operation, and the epoxy must be able to either withstand the resulting forces or flex with them. The temperature also varies from one part of the coil to another, coolest at the water inlets and hottest at the outlets, producing internal stresses that can lead to epoxy cracks and insulation failures. Volumes of epoxy without fiberglass or other filler are especially vulnerable to cracking, so the coil and mold are designed to avoid any space with a dimension greater than 1.5 mm (1/16 in).

Because the potential of a multipole of order  $n$  varies as  $a^n$ , higher order multipoles and miniature magnets require higher current density. Also, specialty magnets, such as extraction current septum magnets, require high current density

in the septum region. If the application will not tolerate low duty cycle, short-pulsed magnets, more extreme designs may be necessary. At the Princeton-Penn Accelerator [11], a septum magnet operated with  $>400 \text{ A/mm}^2$  using 1000 psi water pressure to cool each 40-inch conductor. More recently, such problems have been avoided by using Lambertson septum magnets which have a steel septum between the bending region and the zero field region.

**AC magnets** The skin depth in copper is about 8 mm at 60 Hz. For conductors above this size, the AC magnetic fields and currents will not penetrate the conductor completely, negating the expected reduction in resistance of a larger cross section. If the magnet configuration calls for the conductor to be in a region of strong field, then the losses can be aggravated by the need for the flux lines to cross the conductor. By avoiding such designs, i.e. locating the conductor in a low field region, and using small conductors, hollow conductor coils can be used up to 15-30 Hz. Eventually the current carried by the coil is too small, complicating the coil and power supply. At this point one attempts to employ a fully transposed cable of small conductors (“Litz wire”). The coil can be cooled by embedding a water carrying tube in the cable. Alternately one can simply operate at lower current density and lower field. There is no “standard” cable for this application, but as noted above, AC magnets have historically operated at lower field than DC or slowly pulsed magnets.

**Quality assurance methods** Particularly in systems with large numbers of magnets, a variety of means are available to minimize the effect of errors in fabrication of magnet parts and of magnets themselves.

(i) Measurements can be made of sample laminations from individual heats or coils of steel. These can include the  $B$ - $H$  curve for high field permeability and coercive force, as well as critical dimensions such as gap. Magnets can then be created from the appropriate mix of laminations to average out the effects of fluctuations in these properties.

(ii) The laminations can be designed to accommodate flipping, or alternation as noted above, to average out asymmetries. The design should incorporate a “witness mark” so that the pattern of alternation is evident. The lamination can also include built-in fiducial surfaces for surveying the magnet in its final location.

Table 1: Antiproton Source Magnets (FNAL)

| Type*                               | D(lg) | D(sm) | Q(lg) | Q(sm) |                |
|-------------------------------------|-------|-------|-------|-------|----------------|
| Strength <sup>†</sup>               | 1.7   | 1.7   | 8.9   | 10    | T, T, T/m, T/m |
| Length*                             | 4.57  | 4.57  | 0.874 | 1.28  | m              |
| 2a                                  | 60.3  | 60.3  | 168.3 | 88.9  | mm             |
| GFR**                               | 254   | 120   | 254   | 120   | mm             |
| Current                             | 1178  | 1178  | 1206  | 235   | A              |
| $\Delta S/S_{rms}^{\dagger\dagger}$ | 5     | 5     | 10    | 15    | $10^{-4}$      |
| $\Delta S/S_{max}^{***}$            | 3     | 25    | 25    | 3     | $10^{-4}$      |

Table 2: Advanced Photon Source Magnets (ANL)

| Type                                | Dipole | Quad | 6-pole |                 |
|-------------------------------------|--------|------|--------|-----------------|
| Strength <sup>†</sup>               | 0.6    | 18.9 | 415    | T, T/m, $T/m^2$ |
| Length*                             | 3.06   | 0.8  | 0.253  | m               |
| 2a                                  | 60     | 80   | 98     | mm              |
| Current                             | 450    | 386  | 160    | A               |
| $\Delta S/S_{rms}^{\dagger\dagger}$ | 4.4    | 1.1  | 3.2    | $10^{-4}$       |
| $\Delta S/S_{max}^{***}$            | 0.4    | 2    | 2.5    | $10^{-4}$       |

Table 3: Main Injector Magnets (FNAL)

| Type                                | Dipole | Quad | 6-pole |                 |
|-------------------------------------|--------|------|--------|-----------------|
| Strength <sup>†</sup>               | 1.73   | 19.6 | 120    | T, T/m, $T/m^2$ |
| Length*                             | 6.10   | 2.93 | 0.495  | m               |
| 2a                                  | 50.8   | 83.4 | 52.8   | mm              |
| Current                             | 9420   | 3630 | 300    | A               |
| $\Delta S/S_{rms}^{\dagger\dagger}$ | 4      | 4    | 18     | $10^{-4}$       |
| $\Delta S/S_{max}^{***}$            | 0.2    | 1    |        | $10^{-4}$       |

\* D=dipole, Q=quad, lg=large, sm=small

† Field or gradient

\* Maximum of several magnetic lengths

\*\* Good field region full width

†† RMS fractional variation in integrated strength on center

\*\*\* Max fractional variation in integrated strength (field or gradient) over good field region

\*\*\*\* RMS fractional error in integrated strength at 25 mm

(iii) The location of individual magnets in the lattice can be assigned based on measurements of the fields of the magnets. The general approach is to attempt to cancel contributions to the appropriate Fourier-Floquet harmonic coefficients of the error fields responsible for the excitation of resonances by individual multipoles. For example the closed orbit error is driven by dipole errors, half integral resonance and amplitude variation by quadrupole errors, etc. Performance improvements by factors of 5-10 can easily be achieved. Several multipoles can be corrected at the same time, but this usually leads to less improvement for each.

**Examples of recent magnets built** Tables 1, 2, and 3 describe the major magnets for three recent

projects [12]. Typical lengths, operating fields, and currents are given, as well as the strength variations. It is customary, as in Tables 2 and 3, to give the rms field error (usually for many multipoles, but here as a maximum) at some reference radius (such as 25 mm) as input to tracking codes. In Table 1, the system was required to store beam at widely different locations (note that the quadrupoles had a good field region of about 1.5 pole tip diameters) so the maximum strength variation is shown as a tolerance, which was met.

## References

- [1] For example, E.L. Hubbard, ed., FERMILAB-TM-0405, “Booster Synchrotron” (1973) and

- G.P. Jackson, et al., FERMILAB-TM-1991, "The Fermilab Recycler Ring Technical Design Report." (1996)
- [2] S. C. Snowdon, "Magnet End Termination," FERMILAB-FN-0184, regarding DESY combined function magnets.
  - [3] M.E. Rose, PR 53 (1938) 715
  - [4] S. C. Snowdon, "On the Calculation of Rose Shims," FERMILAB-TM-0710
  - [5] Programs available from LANL Accelerator Code Group; LA-UR-90-1766
  - [6] Available from Vector Fields, Inc., 1700 N. Farnsworth Ave., Aurora, IL 60505
  - [7] Courtesy B.C. Brown, FNAL
  - [8] C. Theisen, Modulus of Elasticity of Steel Laminations, Feb. 1978, BNL
  - [9] In most magnets the flux lines turn through 180° or more in the iron thus negating advantages of orientation. "Feathering" or interlacing laminations in corner regions is used in transformers where there is no serious precision requirement but is too expensive for use in accelerator magnets.
  - [10] J. Gunn, UCRL Design Note (1973)
  - [11] J.L. Kirchgessner, private communication
  - [12] Courtesy B.C. Brown, D.J. Harding, FNAL, and S. Kim, ANL

### 7.2.2 Consequences of Saturation of High Permeability Material

*K. Halbach (deceased) and R. Schlueter, LBNL*

To design and understand the most important properties of magnet systems, it is usually adequate to assume that the permeable material (hereafter referred to as "iron") has permeability  $\mu = \infty$ . When dealing with finite  $\mu$  effects it is assumed that the iron is magnetically isotropic and has a  $B(H)$  curve with these typical properties: a saturation induction of 2 T, and intersections with the  $B$  and  $H$  axes at  $B_r = 0.6\text{--}0.8$  T, and at  $-0.0001$  T =  $-H_c$ . When one considers the material in the region  $H = 0, B = B_r$ , and applies an  $H$  in the direction perpendicular to the existing  $B$ , it is obvious that  $B$  and  $H$  are not parallel to each other any more, i.e. the material is highly anisotropic there. We ignore this in our discussion since this anisotropy is rarely of great importance. To explain the various effects caused by finite permeability we use one quarter of an up-down, right-left symmetrical window frame dipole magnet (Fig.1), with the coil indicated by the cross at

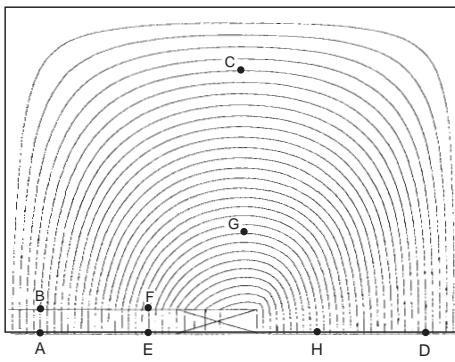


Figure 1: Field lines in a dipole.

the right hand side of the good field region. The two main properties of magnets that are affected by the finite  $\mu$  of iron are the overall field level and the field distribution in the good field region. Finite  $\mu$  has also an effect on small perturbations in magnets (discussed later).

**Field level** The Ampere integral (AI) (defined as  $\int \vec{H} \cdot d\vec{l}$ ) along the path A-B-C-D-A tells us that the field level in the vacuum region is given by the Ampere turns in the coil minus the AI along that field line in the iron. This loss of excitation can present a major problem when many supposedly identical magnets are run by one power supply: assuming a 1% loss of excitation from this source, and assuming a 10% accuracy of that loss due to variability of iron properties between magnets, significant orbit deviations can result.

Applying this to a magnet with the power supply turned off, and using the fact that at the iron-vacuum interface the field  $B$  is continuous, leads to the conclusion that with no coil excitation, the field in the gap will be of the order  $H_c$  times the typical ratio in that magnet of the length of a field line in the iron, divided by the good field region gap. This field level is a good indicator of difficulties one is likely to encounter at low magnet excitation.

**Field distribution** Applying the logic used in the field level consideration to two paths, namely A-B-C-D-A and E-F-G-H-E, leads to the conclusion that the midplane field distribution errors are given by the differences in the AI along the two field lines in the iron. It is often argued that because of the much greater length of the field line B-C-D than the field line F-G-H, the

field increases in the midplane when one moves away from the vertical symmetry line. While true, this logic is incomplete: When one moves by  $dR$  in the direction perpendicular to a (curved) field line that has a radius of curvature  $R_a$ ,  $H$  will increase toward the center of curvature by  $dH = HdR/R_a$ . It is impossible to decide which of the two AI in the iron is larger: While path F-G-H is shorter than the other,  $H$  is larger everywhere along that path than on B-C-D. To make a logically sound argument, we recall that if a field line forms angles  $\alpha_1$  and  $\alpha_2$  with the normal to an interface, with the media on the two sides of the interface characterized by the permeabilities  $\mu_1$  and  $\mu_2$ , these quantities are related through

$$\tan \alpha_1/\mu_1 = \tan \alpha_2/\mu_2 \quad (1)$$

From this follows three simple facts:

(i) If  $\mu_1 \ll \mu_2$ ,  $\alpha_1$  will be “invisible” on a field line plot even when  $\alpha_2$  is as large as  $45^\circ$ .

(ii) Since the tangential field component of  $H$  at an interface is continuous, a “large”  $\alpha_2$  means that  $H$  will have a significant tangential component at the vacuum-iron interface, i.e. finite  $\mu$  will cause field errors, and a field line plot as shown in Fig.1 is an excellent way to judge the sensitivity of a design to saturation of the iron. Taking the AI around the loop A-B-F-E-A shows that the difference in the fields in the good field regions is controlled by the AI along the vacuum-iron interface. Execution of this analysis allows precise estimates of these field errors and shows in particular that these field errors are inversely proportional to the gap. Specifically, assuming  $\alpha_2$  increases linearly from  $\alpha_2 | (x=0)$  to  $\alpha_2 | \text{mid}(x=x_o) = \alpha_{2o}$  then

$$\frac{\Delta B_y(x)}{B_y | (x=0)} = \left( \frac{1}{2} \right) \left( \frac{x^2}{x_0 h} \right) \left( \frac{\alpha_{2o}}{\mu_2} \right) \quad (2)$$

(iii) It is generally extremely dangerous to make any judgments based on the comparison of the AI taken along different field lines.

**Small perturbation effects** When looking at the consequences of small perturbations of a magnet, the nonlinear isotropic iron has to be represented, for the purpose of assessing the effects of small perturbations, by a linear anisotropic medium with the permeabilities

$$\mu_{\parallel} = dB/dH, \quad \mu_{\perp} = B/H \quad (3)$$

in the directions parallel and perpendicular to the local field.

### 7.2.3 Special Topics in Magnetics

#### 7.2.3.1 Properties of 3D vacuum fields integrating along a straight line *K. Halbach (deceased) and R. Schlueter, LBNL*

If  $B_z(x, y, z_1) = B_z(x, y, z_2)$ , and  $L$  designates a convenient reference length, such as from one zero field region to another (e.g. upstream and downstream of a magnet), then

$$\begin{aligned} \langle B_x \rangle &= \langle B_x(x, y) \rangle = \frac{1}{L} \int_{z_1}^{z_2} B_x(x, y, z) dz, \\ \langle B_y \rangle &= \langle B_y(x, y) \rangle = \frac{1}{L} \int_{z_1}^{z_2} B_y(x, y, z) dz \quad (1) \end{aligned}$$

satisfy the “normal” 2-D differential equations for 2-D vacuum fields. This means specifically that, with  $Z = x + iy$ ,  $B = B_x + iB_y$  is an analytical function of  $Z^* = x - iy$ ; or equivalently,  $B^*$  is an analytical function of  $Z$ . Deriving these 2-D fields from the  $z$ -component  $A(x, y)$  of the vector potential and the scalar potential  $V(x, y)$  (including a multiplicative factor  $\mu_0$  to “fix” the MKS dimensions), the complex potential  $F = A + iV$  is also an analytical function of  $Z$ , and  $B^*$  is obtained from  $F$  through

$$B^* = idF/dZ \quad (2)$$

The remarkable simplicity of the mathematics of 2-D fields combined with the very useful information 2-D fields give about essential aspects of 3-D fields is responsible for the extensive use of 2-D fields.

#### 7.2.3.2 Pole width necessary to obtain desired field quality in a 2-D magnet *K. Halbach (deceased) and R. Schlueter, LBNL*

Addressing first the design of a dipole, Fig.1 shows schematically a cross section of the upper half plane of a symmetrical dipole, with only the left half of the good field region and the pole shown. There the pole extends by  $d$  beyond the good field region, and in order to get the field quality  $e = dB/B$  in a magnet with half gap  $h$ , the magnitude of  $\nu = d/h$  to obtain  $e$  depends on whether or not the pole face is shaped in order to reduce the pole width for a given  $e$ . Assuming  $\mu = \infty$  iron, these are the pole overhang guidelines:

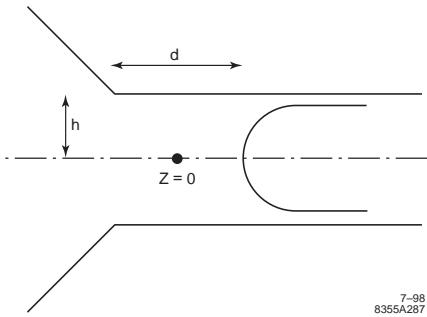


Figure 1: Cross section of a symmetric dipole.

|            |   |
|------------|---|
| With shims | $\begin{cases} \nu = -.14 \ln(e) - .25 \\ e = \exp[-7.14(\nu + .25)] \end{cases}$ |
| No shims   | $\begin{cases} \nu = -.36 \ln(e) - .9 \\ e = \exp[-2.78(\nu + .9)] \end{cases}$   |

The given field quality with a shimmed pole can, in general, only be achieved with a systematic magnet shimming design program like MIRT [2]; trial and error rarely achieves the indicated performance. The shimmed performance will in most cases be achieved for the design field level even when  $\mu$  is finite unless it is too small; however it is difficult to achieve the design field quality over a significant range of field levels. The flat pole field quality will be less than the given guidelines the smaller the permeability.

Specification, measurement, design, and field quality check of fields should not be done inside of the good field region, but on its boundary. The reason is this: The Cartesian field components can be obtained by first derivatives of the scalar potential. Since the scalar potential satisfies the Laplace equation, the desired fields and the actually produced fields, as well as the differences between these fields (i.e. the field errors), also satisfy the Laplace equation. This means that the largest and smallest differences between desired and achieved fields must be on the boundary of the good field region, never inside it. These facts apply to 3-D fields and 2-D fields, and in the latter case also when the design is executed in, as suggested below, conformally mapped geometry. The reason for formulating the overhang guidelines for a semi-infinite magnet is that field perturbations decay exponentially with the distance from the source of the perturbation. Formulating the perturbation fields for locations to the right of

the perturbation source, they are given by

$$dB_z^* = \sum_{n=1} \left[ i b_n e^{-n\pi Z/h} + c_n e^{-(n-1/2)\pi Z/h} \right] \quad (3)$$

making the considerations for the two ends of the pole independent of each other. In Eq.(3),  $b_n$  and  $c_n$  are real and describe the magnitude of error fields that are, respectively, perpendicular and parallel to the midplane in the midplane, and are perpendicular to the poles at the poles at  $y = \pm h$ . Eq.(3) is important for many considerations and is the reason why one can make often simple and accurate approximations for seemingly difficult problems. The determination of the width and shape of the pole(s) of a non-dipole can be reduced to that task for a dipole with the following procedure: Assuming that, as is customary, the field is specified by  $B_x(x, 0) = 0$ ,  $B_y(x, 0) = g(x)$  in the midplane, it follows directly that the ideal 2-D complex field is given by  $B^* = ig(Z) = idF/dZ$ . The conformal map  $W(Z)$  that maps the field shaping entities (like coils, iron poles, infinite conductivity surfaces for pulsed systems, etc.) of an ideal non-dipole into those of an ideal dipole is given by any constant times  $F(Z)$ , leading to

$$W(Z) = C \int g(Z) dZ \quad (4)$$

The design is then accomplished by mapping with  $W(Z)$  both the good field region, vacuum chamber, and whatever else of interest, from the  $Z$  plane to the  $W$  plane, designing there the essentials of the pole with the help of the guidelines given above, and then mapping it back into the  $Z$  plane. When executing this design, it is profitable to take advantage of magnet symmetries that usually mean that one has to design only a small fraction of the total magnet. Keep in mind also that relative field errors in the  $Z$  plane are the same as in the  $W$  plane, i.e.  $dB_Z/B_Z = dB_W/B_W$ .

### 7.2.3.3 Eddy currents

*K. Halbach (deceased) and R. Schlueter,  
LBNL*

Of the many eddy current effects that can cause problems (or, sometimes, can be used advantageously) only two will be discussed here. They stand out because they are poorly understood by many designers and users of magnets. The first is the simple concept of skin depth, and what it may mean when the eddy currents are induced in

a high  $\mu$  material. We are referring specifically to fast, usually small amplitude, field changes. The second example is a solid core dipole magnet in which eddy currents may drive some of the iron into saturation, contrary to what is sometimes expected on the basis of “Lenz’s law”.

**Skin depth related problems** A  $x$ -independent sinusoidal field  $H_x$  parallel to the surface of a conductor propagates into the conductor ( $y > 0$ ) according to  $H_x(y) = H_x e^{-ky}$  with  $D_1 = 1/k = 1/\sqrt{\sigma_c \mu_0 \mu i \omega}$  representing the conventional skin depth. When  $H_x$  depends on  $x$ , flux conservation leads to the following boundary condition at the interface between vacuum and the conductor:

$$H_y(x) = D_2 \frac{\partial H_x(x)}{\partial x} \quad (5)$$

with  $D_2 = \mu D_1$ . If the conductor is significantly thicker than  $D_1$ , the fields in the vacuum region are, through this boundary condition, completely controlled by  $D_2$ , no matter what  $D_1$  is, making what is commonly called “the skin depth” ( $D_1$ ) totally irrelevant there. The extreme cases  $\mu = \infty$  and  $\sigma_c = \infty$  give the boundary conditions for extreme conditions of infinitely permeable iron ( $\partial H_x / \partial x = 0, H_x = 0$ ) and perfectly superconducting material ( $H_y = 0$ ).

**Eddy current-caused fields in the iron of a solid core magnet** Fig.2 represents a schematic of one quarter of a solid core 2-D dipole magnet with fields perpendicular to the horizontal drawing boundary and parallel to the right boundary, and also showing 1/2 of a coil. As with all air gap magnets, the flux carried by the iron yoke is controlled to the % accuracy level by the gap and the current in the coil. When changing the current, eddy currents will produce field changes in the iron. Since the flux carried by the iron is essentially independent of the eddy currents, this means that while the field will be decreased in some places, it will increase elsewhere.

To get a qualitative and semi-quantitative idea of what happens, we describe only the events for the simple case of a “slow” (turn on in minutes) “large amplitude” (turning on or off to/at the 1 T field level) linear current ramp. For simplicity, it is assumed that  $\mu$  is constant, and the fields are calculated for the region where the midplane of the magnet intersects the yoke. Describing the field level there as a function of time for the case of the linear ramp by  $B_0(t)$ , the field at location characterized by  $e = x/d$  is given, after transients

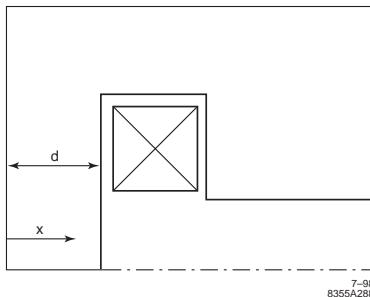


Figure 2: Schematic of a solid core dipole.

have decayed, by

$$B(e, t) = B_0(t) + \dot{B}_0(t)\tau(3e^2 - 1)/6 \quad (6)$$

with  $\tau = \sigma_c \mu_0 \mu d^2$ . Using (characteristic for 1010 steel):  $\sigma_c \mu_0 \mu = 30,000$ ,  $d = 0.5$  m,  $e = 1$  (corresponding to the iron next to the vacuum and coil region),  $\tau = 7500$  s, and one obtains at  $x = d$  for a linear ramp that turns the magnet to 1 T in 10 minutes,  $dB = \dot{B}_0(t)\tau/3 = 4$  T. While this is obviously nonsense, it indicates that in reality the iron in that region will saturate at a level of about 2 T, when  $\mu$  will go down. The situation is even more aggravating when one turns a 1 T magnet off: the iron in the region next to the vacuum region will be saturated in the opposite direction, and the early part of that process will occur very fast since the iron will first be driven to a region of the  $B(H)$  curve where  $\mu$  will increase. One consequence is that the iron will have unexpected  $\mu$  even after all eddy currents have decayed, because the iron was driven temporarily into saturation [3]. Essentially the same phenomenon occurs in laminations. In contrast to the solid core magnet, it is easy to evaluate every detail accurately for any  $B(H)$  curve and excitation pattern since in that case one is dealing (in good approximation) with the 1-D diffusion equation with a nonlinear medium, a simple problem. It may be worthwhile to point out that one can use this behavior of a laminated magnet to advantage: when, in an ac magnet, one goes down to the lowest field level, one can force the laminations at their edges (and that is where it counts) to be driven into the third quadrant of the  $B(H)$  curve, using on the following field upswing a more favorable part of the  $B(H)$  curve (namely intersecting the  $B = 0$  level for  $H > 0$ ), without needing a bipolar power supply.

### 7.2.3.4 Power dissipation in the dipole coils of a storage ring with iron poles

*K. Halbach (deceased) and R. Schlueter, LBNL*

Notation (MKS units):  $P$  = power,  $g$  = gap,  $p$  and  $e$  = momentum and charge of particles,  $j$  and  $\rho_r$  = current density and resistivity in coil material,  $w/l$  = width to length ratio of coils.

$$P = g\left(\frac{p}{e}\right)j\rho_r \frac{4\pi}{\mu_0}(1 + \frac{w}{l}) \quad (7)$$

As is often the case, it is of great importance to appreciate which quantities do not appear in the expression for  $P$ :  $P$  is independent of the ring circumference and the field level, at least as long as the iron is not very strongly saturated.

### References

- [1] A. Asner, P. Bossard, Ch. Iselin, Proc. Int. Symp. Magnet Technology (1965)
- [2] K. Halbach, Proc. Int. Conf. Magnet Technology (1967)
- [3] K. Halbach, NIM 107 (1974) 529

### 7.2.3.5 Alpha magnet

*H. Wiedemann, Stanford U./SSRL*

Alpha magnets [1] are used for off-axis injection into a linear accelerator or for bunch compression. The magnet is half of a quadrupole, sliced vertically with a ferromagnetic mirror plate installed at the midplane. Particle injected at an angle of  $\theta_\alpha = 40.70991^\circ$  with respect to the magnet axis and at the mid-point of the magnet follow an  $\alpha$ -like path and exit again at the entrance point with a total deflection of  $2\theta_\alpha$  irrespective of the particle energy (see Fig.3). The length  $s_\alpha$  and apex  $\hat{u}$  of the  $\alpha$ -path are given, respectively by

$$s_\alpha[\text{m}] = 0.191652 \sqrt{\frac{\beta\gamma}{g[\text{T/m}]}} \quad (8)$$

$$\hat{u}[\text{m}] = 0.0750498 \sqrt{\frac{\beta\gamma}{g[\text{T/m}]}}$$

where  $\beta = v/c$  and  $\gamma = E/mc^2$ . The path length is energy dependent and makes therefore an  $\alpha$ -magnet suitable for efficient bunch compression.

### References

- [1] H.A. Enge, RSI 34 (1963) 385; M. Borland, PhD thesis, SLAC-R-0402 (1993)

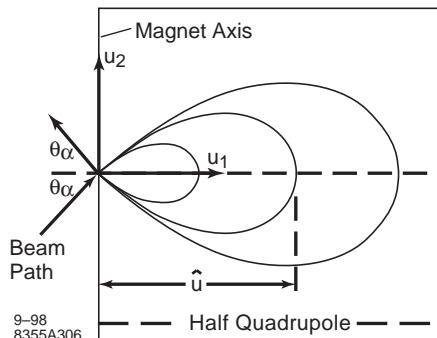


Figure 3: Schematics of an alpha magnet.

### 7.2.4 Cosθ Superconducting Magnets

*P. Schmüser, U. Hamburg*

**General Properties** Cosθ SC accelerator magnets are operated at fields between 3.5 and 8.5 T, far beyond the saturation magnetization of iron. Much useful information will be found in [1, 2, 3, 4, 5]. The magnets are current-dominated and extreme precision in coil geometry (typically 20  $\mu\text{m}$ ) is needed to achieve a field homogeneity in the  $10^{-4}$  range. Precision stamping is used to produce the tooling for coil fabrication and the clamps for collaring the finished coil. The huge magnetic forces acting on the coil at full excitation ( $\sim 10^6 \text{ N/m}$  at 5 T) are taken over by the clamps and/or the soft-iron yoke. The iron yoke is basically a hollow cylinder surrounding the coil which enhances the central field but screens the fringe field outside the magnet. All existing accelerator magnets are made from multifilamentary niobium-titanium superconductor. The maximum operating field is 6.5 T with normal helium cooling at 4.3 K and 8.5 T with superfluid helium cooling at 1.9 K. Niobium-tin prototype dipoles have exceeded 14 T [6] but the material is very brittle and requires extended heat treatment after coil winding.

In the 2-4 K regime liquid helium is the only substance with an appreciable heat capacity. Helium porosity of the coils is hence an important prerequisite to achieve a stability against quenches due to conductor motion or beam losses. In case of a quench the current has to be bypassed around the quenched magnet using diodes or thyristors. Each magnet in the accelerator has to be able to absorb its own stored field energy (up to several MJ).

A particular drawback of SC magnets are the persistent magnetization currents in the NbTi filaments which seriously affect field quality at low excitation and necessitate more sophisticated correction schemes than in normal accelerators.

**Field description** (See Sec.6.17.1) The complex magnetic field can be expanded in a multipole series in Cartesian coordinates,

$$B_y + iB_x = B_{\text{ref}} \sum_{n=1}^{\infty} (b_n + ia_n) \left( \frac{x+iy}{r_0} \right)^{n-1} \quad (1)$$

where  $r_0$  is a reference radius which should exceed the largest beam deviation from the axis ( $r_0 = 25$  mm for Tevatron and HERA, 17 mm for LHC). The  $b_n$  are called the normal multipole coefficients, the  $a_n$  are the skew coefficients. With the above choice of  $B_{\text{ref}}$  the main coefficient is normalized to unity:  $b_1 = 1$  in a dipole,  $b_2 = 1$  in a quadrupole, etc. (In US practice, the multipole series begins with  $n = 0$ .) The remaining coefficients are small for a good magnet,  $|a_n|, |b_n| < 10^{-4}$ . A coil with dipole symmetry possesses only normal multipoles of odd order (the so-called “allowed” multipoles) while the “forbidden” poles ( $b_n$  of even order and all  $a_n$ ) are absent.

A pure multipole field, containing just the single order  $n = m$ , is obtained for a current distribution of the form

$$I(\theta) = I_0 \cos(m\theta) \quad (2)$$

In practical magnets, the ideal  $\cos \theta$  current distribution of a dipole ( $\cos 2\theta$  for a quadrupole) is approximated by current shells and/or current blocks, see Fig.1a [17]. In a single-shell dipole coil of average radius  $a$  and limiting angle  $\phi_l$  the multipole coefficients  $b_n$  are

$$b_n = \frac{1}{n} \left( \frac{r_0}{a} \right)^{n-1} \frac{\sin(n\phi_l)}{\sin(\phi_l)}; \quad n = 3, 5, 7, \dots \quad (3)$$

Choosing  $\phi_l = 60^\circ$ , the sextupole  $b_3$  vanishes but the decapole,  $b_5$  is two orders of magnitude larger than tolerable. Using two shells, as in Fig.1b, with internal spacers or a sufficient number of current blocks, all higher multipoles can be suppressed below  $10^{-4}$ .

**Influence of the iron yoke** Superconducting dipoles and quadrupole magnets are surrounded with an iron yoke with a cylindrical inner bore to reduce the fringe field. The yoke contribution to the central field can be computed with the image current method provided the iron is not saturated and the permeability  $\mu$  is uniform. For a

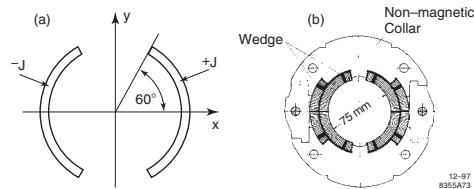


Figure 1: One and two shell dipoles.

single-shell coil the yoke contribution to the central dipole field is

$$(B_1)_{\text{yoke}} / (B_1)_{\text{coil}} = (a/R_y)^2 \quad (4)$$

where  $a$  is the average coil radius and  $R_y$  the inner yoke radius. The relative iron contribution to the dipole field depends on the proximity between coil and yoke and amounts to 10% in the Tevatron, 22% in the HERA and LHC, and 35% in the RHIC magnets. For higher multipoles  $n = 3, 5, 7, \dots$  the iron contribution is considerably smaller,

$$(b_n)_{\text{yoke}} / (b_n)_{\text{coil}} = (a^2/R_y^2)^n \quad (5)$$

An unsaturated yoke does not create “unallowed” multipoles but may change the allowed poles like  $b_3$  since in a two layer coil the mirror images of the inner and outer coils are inverted. The limiting angles of the coil shells are adjusted in such a manner that the sextupole vanishes when the coil is mounted in the yoke. The LHC twin aperture dipole, Fig.2, possesses a small normal quadrupole  $b_2$  because of the left-right asymmetry.

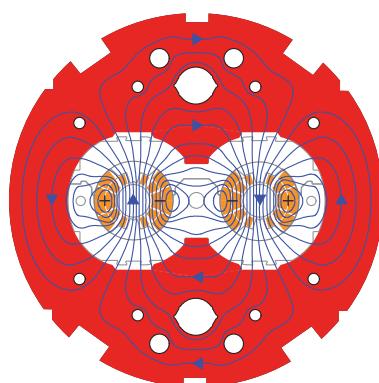


Figure 2: LHC twin bore dipole with computed field lines (courtesy S. Russenschuck, CERN).

Saturation of yoke and end field The image current method fails when the yoke saturates. Finite-element programs are needed to compute the field pattern. Saturation effects are almost negligible in so-called “warm-iron” magnets, featuring an iron yoke outside the cryostat, but may become large in “cold-iron” magnets with the yoke directly surrounding the insulated coil. By punching suitable holes into the iron-yoke laminations the saturation-induced higher multipoles can be strongly suppressed.

The field computation in the coil head region relies on numerical methods. The windings are usually spread out by insulating spacers which provide mechanical stability and prevent a field enhancement in the coil ends.

### Mechanical accuracies and magnetic forces

Mechanical tolerances A good approximation of the ideal  $\cos \theta$  current distribution of a dipole can be realized by a suitable arrangement of current shells and/or blocks. In real magnets the manufacturing tolerances are the dominating uncertainty. If for example the limiting angle  $\phi_l$  of a single-shell coil differs from the ideal value of  $60^\circ$  the sextupole coefficient is no longer zero,

$$b_3 = \frac{1}{3} \left( \frac{r_0}{a} \right)^2 \frac{\sin(180^\circ + 3\delta\phi_l)}{\sin(60^\circ + \delta\phi_l)} \quad (6)$$

where  $\delta\phi_l$  is the angular error. The condition that  $|b_3| < 10^{-4}$  requires  $|\delta\phi_l| < 0.25$  mrad. From an analysis of all conceivable geometry distortions, one concludes that mechanical accuracies in the  $20 \mu\text{m}$  range are needed to satisfy the field homogeneity requirements of a large storage ring. Such narrow tolerances are difficult to accomplish by conventional machining but can be realized using precision-stamped laminations to assemble the tooling for coil winding and baking and also for the collars which clamp the finished coil.

The centering of a dipole coil with respect to the cylindrical iron yoke must be accurate to  $0.1 \text{ mm}$  to avoid the creation of quadrupole terms. This is easily accomplished in a cold-iron magnet by means of tongue-and-groove interlocks between collar and yoke, while in magnets whose yoke is outside the cryostat, many support planes and a precise alignment are needed.

Multipole measurements provide the most sensitive tests on the precision of the coil cross section and of the placement of the current conductors. Fig.3a shows the normal multiple coefficients of 200 HERA dipoles. Most of the coefficients are well within the limits of  $\pm 0.5 \times 10^{-4}$

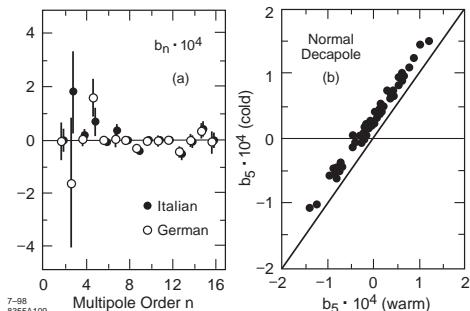


Figure 3: (a)  $b_n$  of HERA dipoles at 4.66 T (800 GeV). Plotted are the average values for 200 magnets with standard deviations. Data averaged over whole magnet including coil ends. (b) Correlations of room temp. and cryogenic measurements of decapole in RHIC dipoles (R. Gupta, BNL). The systematic offset is caused by the fact that the cryogenic data were taken at high current where iron saturation plays a role.

used in the particle tracking program for determining the dynamic aperture (Sec.2.3.9) of the HERA storage ring. An exception are the normal sextupole  $b_3$  which is particularly sensitive to slight changes in the limiting angles of the coil shells, and the skew quadrupole  $a_2$  which may arise from an up-down asymmetry of the coil. The dipoles have been “sorted” in the accelerator in order to minimize the sextupole effects on the beam. Since cryogenic measurements are very time consuming there is considerable interest whether measurements in the normal-conducting coils at room temperature are sufficiently accurate for determining the field quality. This is indeed the case as can be seen from Fig.3b. The good “warm-cold” correlation allowed to restrict cryogenic measurements to a 10% fraction of the RHIC magnets without sacrificing field quality. At LHC 15 – 20% of the magnets were measured cryogenically, the field quality conformed to beam optics requirements [7].

The field integral of the dipoles and quadrupoles can be measured with an accuracy in the  $10^{-4}$  range. The rms variation in the HERA magnets is about  $5 \times 10^{-4}$ . The axis of the RHIC quadrupoles and sextupoles is determined with an rms precision of  $0.05 \text{ mm}$  using a ferrofluidic cell viewed with polarized light [8].

Magnetic forces The two halves of a dipole coil repel each other with a force of about  $10^6$

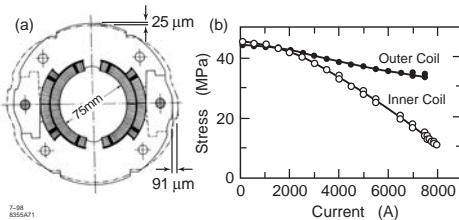


Figure 4: (a) Calculated deformation of the collared HERA coil in the superconducting state at 6T (G. Meyer, DESY). The collar material is AlMg4.5Mn (4.5% Mg, 0.4-1% Mn)(G35) having a yield strength of 270 MPa at room temp. Max. stress is 150 MPa. (b) Measured pre-stress in the inner and outer coil shell of an SSC model dipole (courtesy C. Goodzeit).

N (100 tons) per meter length at a field of 5 T. In addition large azimuthal forces act on the conductors near the limiting angles of the coil shells which are directed towards the median plane. A conductor motion away from the end stops provided by the collars must be strictly avoided since the accompanying frictional heat triggers premature quenches. For this reason the coils are mechanically precompressed with a pre-stress exceeding the maximum magnetic pressure. The collars must be sufficiently stiff to apply the required pre-stress (60 MPa in the HERA magnets) and to sustain the large horizontal Lorentz forces. Fig.4a shows the computed deformation of the aluminum-collared HERA dipole coil in the SC state at 6 T.

In the LHC magnets a higher pre-stress of 100 MPa is needed and the iron yoke serves as an additional support structure. The precompression in dipole coils has been measured at BNL using strain gauges. The pre-stress in a 1.8-m-long SSC model dipole is plotted in Fig.4b as a function of current. Both coil shells retain a positive pre-stress far beyond the nominal current of 6600 A. In the coil heads Lorentz forces of about 15 tons at 5 T act in longitudinal direction and tend to lengthen the coil. The forces are taken up by stainless steel end plates welded to the stainless steel LHe container.

**Persistent currents** Persistent magnetization currents in the filaments of the SC cable generate all multipoles which are allowed by coil symmetry:  $b_1, b_3, b_5, \dots$  in a dipole,  $b_2, b_6, b_{10}, \dots$  in a quadrupole. Due to flux pinning, the multipole fields exhibit a pronounced hysteresis with

opposite signs for increasing and decreasing main field, respectively.

The computation of persistent-current effects is based on the critical state model [9] whose basic assumption is that a hard superconductor tries to expel external field changes by generating bipolar shielding currents with the critical current density  $J_c = J_c(B, T)$ . A filament subjected to a homogeneous external field, which is raised from zero to a value  $B_{ext}$  sets up a  $\cos\theta$  like shielding current [3]. The current- and field-free region can be approximated by an ellipse with large half axis,  $a = r_f$  (filament radius), small half axis  $b$  and eccentricity  $\epsilon = \sqrt{1 - (b/a)^2}$ . The shielding field is [3]

$$B_{sh} = -\frac{2\mu_0 J_c r_f}{\pi} \left( 1 - \sqrt{1 - \epsilon^2} \frac{\sin^{-1} \epsilon}{\epsilon} \right) \quad (7)$$

The highest field which can be shielded from the interior of the filament is called the “penetrating” field,  $B_p$ , and is obtained for an ellipse shrunk to a line:  $B_p = 2\mu_0 J_c r_f / \pi$ . External fields above  $B_p$  (about 0.1 T for NbTi) lead to a non-vanishing field throughout the filament. If the external field is decreased again, persistent currents with opposite polarity are superimposed. The magnetization of the filament, defined as the magnetic moment per unit volume, is

$$M = -\frac{4}{3\pi} J_c r_f \epsilon^2 \quad (8)$$

$M$  assumes its peak value for the fully penetrated filament ( $\epsilon = 1$ ),

$$|\hat{M}| = (4/3\pi) J_c r_f \quad (9)$$

which decreases proportional to the critical current density  $J_c(B_{ext}, T)$  when the external field is raised beyond the penetrating field. The field-generating transport current (density  $J_t$ ) is taken into account by a correction factor  $[1 - (J_t/J_c)^2]$ . The correction is negligible at low field but leads to a significant reduction of filament magnetization at high excitation of the magnet.

The dependence of filament magnetization on the external field is characterized by a hysteresis [2] resembling that of iron but with opposite sign – the superconductor is diamagnetic. The persistent-current multipoles in a dipole are conveniently computed starting with four symmetrically arranged filaments in the coil in which bipolar currents have been induced by the increasing

main field (see Fig.5a). The vector potential of the four current pairs with a magnetic moment per unit length of  $m = \pi r_f^2 M$  can be expressed as

$$A^{\text{Pair}}(r, \theta) = -\frac{2\mu_0 m}{\pi R} \sum_{n=1,3,\dots} \left(\frac{r}{R}\right)^n \times \cos(n\theta) \cos(n\phi + \alpha) \quad (10)$$

$R$  and  $\alpha$  are defined in Fig.5a. Eq.(10) has to be summed over all filaments in the first quadrant of the coil.

Eq.(10) shows that the persistent currents generate exactly the same “allowed” multipoles as the transport current, namely the normal multipoles  $b_n$ ,  $n = 1, 3, 5, \dots$ . “Unallowed” poles arise only from symmetry distortions, e.g. when the critical current density of the superconductor is not uniform over the coil cross section. In a quadrupole coil the persistent currents generate only the allowed normal multipoles  $b_2, b_6, b_{10}, \dots$ . Sextupole data from 315 HERA dipoles are found in perfect agreement with the model calculations (Fig.5b) [10]. The persistent-current contribution to the main dipole field (quadrupole gradient) is in the same order as the sextupole (12-pole) and exhibits a similar hysteresis.

For a given superconductor fraction in the cable the number of filaments scales with  $r_f^{-2}$ . Hence the multipoles are proportional to the filament radius which therefore should be chosen as small as technically feasible. A typical value is 3  $\mu\text{m}$  in recent magnets.

Time dependence and longitudinal periodicity of persistent current effects The persistent-current fields exhibit an almost logarithmic time dependence (Fig.6a), which can be partly attributed to thermally-activated flux creep in the superconductor but, to a larger extent, to a complicated interaction between the persistent currents in the superconductor and conventional eddy currents in the copper matrix of the cable.

The measured decay rates vary strongly from magnet to magnet and depend both on magnetic field history as well as on cable properties like interstrand resistance. A quantitative mathematical model is not available. The sextupole component in a dipole as well as other multipoles exhibit a periodic pattern along the axis of the magnet which becomes particularly pronounced at high fields (Fig.6b) [11]. The most likely explanation is a current imbalance among the strands of the Rutherford-type cable. The oscillating multipole fields has a negligible effect on the beam

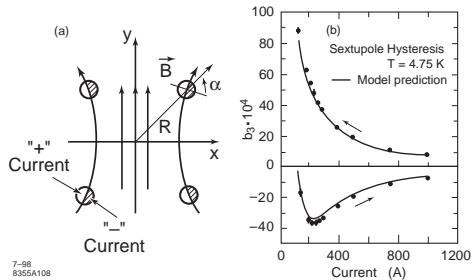


Figure 5: (a) Magnetization currents in four symmetric filaments of a dipole coil. (b) The averaged sextupole coefficients with rms errors from 315 HERA dipoles as a function of current. The curves represent the model predictions. The ramp directions of the current is indicated by the arrows. Before starting the measurements, a field cycle ( $0.05$  T  $\rightarrow$   $5.8$  T  $\rightarrow$   $0.05$  T  $\rightarrow$   $0.23$  T) was performed to establish a well defined initial superconductor magnetization.

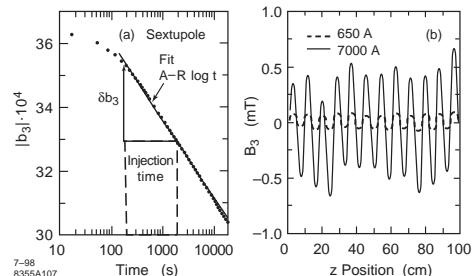


Figure 6: (a) Time dependence of the persistent current sextupole in a HERA dipole at a field of  $0.227$  T. (b) Longitudinal periodicity of sextupole in an SSC dipole at  $650$  A and  $7000$  A.

emittance since the wavelength  $\ll$  the betatron wavelength.

**Eddy Current Effects** In a Rutherford cable, the interstrand resistance  $R_c$  at the cross-over points may be as low as a few  $\mu\Omega$  for strands with bare copper surface. The two layers of strands form an arrangement of loops in which eddy currents are induced during magnetic field ramps. The generated heat per unit length is [12]

$$G \approx (w^2 l_p N^2 / 120 R_c) \dot{B}^2 \quad (11)$$

where  $w$  is the cable width,  $N$  the number of strands and  $l_p$  the cable pitch length. The small-loop cable eddy currents do not obey any coil symmetry because  $R_c$  varies from turn to turn.

Therefore all allowed and unallowed multipoles appear during a ramp of the magnetic field with a magnitude  $\propto \dot{B}$ . Ramp-rate dependent multipoles are correlated with ac losses and, in severe cases, even with a reduction of quench current at high ramp rates. Control of interstrand resistance is thus an important aspect of cable production. At LHC the strands are coated with an oxidized Ag-Sn layer, and the specified interstrand resistance is  $R_c = 15\mu\Omega$  for the inner dipole coil layer and  $40\mu\Omega$  for the outer layer.

A magnetic field cycle in the accelerator is accompanied with heat transfer to the liquid helium due to hysteretic losses in the superconductor (reversal of magnetization) and eddy currents among the strands of the cable. In a HERA dipole the hysteretic loss  $\sim 400$  J per cycle  $0.23\text{ T} \rightarrow 4.7\text{ T} \rightarrow 0.23\text{ T}$  while the eddy-current loss  $\sim 75$  J at a ramp rate of  $\dot{B} = 0.01\text{ T/s}$ .

The small-loop eddy currents decay within less than a second. Another type of eddy currents is due to a variation of  $B$  along the cable. In particular the coil ends contribute here. These “boundary-induced” eddy currents [13] have decay times of many hours or even days and lead to a non-uniform distribution of current among the strands in the cable. They are the physical origin of the longitudinal periodicity (Fig.6b) and are probably also responsible for the large variations in persistent-current decay rates. Quantitative predictions are not available because of too many unknown parameters.

**Transmission line characteristics of a string of magnets** A SC dipole cannot be considered as a pure inductance, in spite of the vanishing resistance of the conductor. The eddy current losses correspond to a frequency-dependent real part of the impedance while the turn-to-turn capacitance modifies the imaginary part. An equivalent circuit of a dipole is shown in Fig.7a. A long string of dipoles has transmission line characteristics. The impedance and phase velocity are  $Z = \sqrt{L'/C'} \approx 900\Omega$  and  $v = 1/\sqrt{L'C'} \sim 10^5\text{ m/s}$ , where  $L'$  and  $C'$  are the inductance and capacitance per unit length.

Transmission line effects in a string of 96 Tevatron magnets are shown in Fig.7b. A pronounced resonance is observed at 100 Hz which is undesirable as it may lead to voltage enhancements during a magnetic field ramp or to dipole field variations along the chain. The resonance can be strongly damped by connecting an external

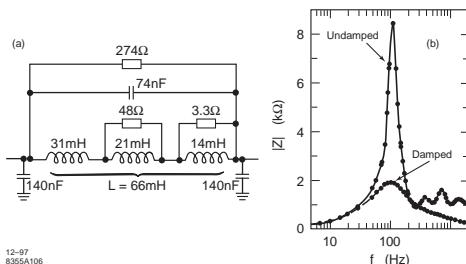


Figure 7: (a) Equivalent circuit of SSC dipole including cable capacitance [14]. (b) Observed resonance behavior in a 96 magnet string in the Tevatron. Shown is the absolute value of the impedance as a function of frequency for the undamped and the damped cases. Curves: model calculation.

resistor ( $100\Omega$ ) across the potential wires of each coil.

**Quenches and magnet protection** Two types of quenches (irreversible transition to normal conductivity) can be distinguished: A “natural” quench occurs when either temperature  $T$ , magnetic field  $B$ , or current density  $J$  are raised beyond their critical values, for instance in magnet tests, while a “disturbance” quench may happen with the nominal working point below the critical parameters if local overheating occurs. The main sources are: conductor motion associated with friction under the influence of large Lorentz forces, and beam loss induced heating. Due to the small heat capacity of the cable at 2-4 K, very little energy is needed to raise  $T$  beyond  $T_c$  (a few milli-Joules per cm $^3$ ). Filling the voids in the cable with LHe greatly enhances the overall heat capacity of the cable and hence its stability. The high heat conductivity of superfluid helium improves the thermal stability of the LHC magnets.

The conservative approach of making the coils cryostable (i.e. the copper in the cable is able to carry the entire current with the available helium cooling) is inadequate for accelerator magnets; it would require a copper-to-superconductor (Cu/SC) ratio of  $> 10$  and make the coils too bulky and expensive. With the usual Cu/SC ratio of  $\sim 1.5$ , accelerator magnets are definitely not cryostable, hence a reliable quench protection scheme is vital.

A “hot spot”, generated by some disturbance, will develop into a quench if the normal-conducting zone is longer than so-called *minimum*

*propagating zone,*

$$l_{\text{mpz}} = \sqrt{2\kappa(T_c - T_0)/\rho_r J^2} \quad (12)$$

Here  $\kappa$  is the average heat conductivity of the composite Cu/NbTi conductor,  $\rho_r$  its average resistivity,  $J$  the current density and  $T_0$  the helium temperature. Typically a normal zone of more than 10 mm length will spread out while a shorter zone will shrink away. The quench propagation velocity in the adiabatic case (helium cooling neglected) is

$$v_{\text{adiab}} = (J/c_p) \sqrt{\rho\kappa/(T_c - T_0)} \quad (13)$$

$c_p$  is the average heat capacity of the cable. The computed velocities are in the order of 10-50 m/s. Contrary to expectation the measured quench propagation velocities in helium-filled coils are even higher, up to 200 m/s in 17-m-long SSC prototype dipoles. The high velocities may be caused by pressure waves in the LHe which propagate a warm front with a speed of about 230 m/s. A high quench propagation velocity is desirable to spread the dissipated energy over a large part of the coil and prevent local overheating. Once a quench has been detected the coil current must be reduced to zero as fast as possible. In a long chain of series-connected magnets the stored magnetic energy of a quenched magnet cannot be dissipated in an external dump resistor but must be absorbed in the normal-conducting part of the coil. The time dependence of the current density after a quench and the maximum temperature in the coil are related by

$$\int_0^\infty J^2(t) dt = \int_{T_0}^{T_{\max}} [c_p(T)/\rho(T) dT] \quad (14)$$

Typical values of  $T_{\max}$  are 150 K which is a moderate level concerning mechanical stresses. Numerical codes exist to simulate quenches [15], [16]. A “typical” quench is described fairly well but the results depend critically on the input assumptions. Measurements on full-size magnets and magnet strings are indispensable to check the validity of the numerical codes and establish the regime of safe operation of the SC accelerator.

Quench detection is based on measurements of resistive voltages. Inductive voltages are eliminated either by subtraction (Tevatron) or by using balanced bridge circuits (HERA). LHC uses both methods. The coil current must decay with a time constant of  $< 1$  s to limit  $T_{\max}$  to safe values below 300 K. The large inductance of a magnet string requires long current decay times ( $> 20$  s

typically) to avoid excessive inductive voltages. Hence a bypass is needed which guides the main current around the quenched coil. In the HERA, RHIC and LHC magnets “cold diodes” inside the LHe container are used for this purpose, in the Tevatron the bypass is provided by thyristors outside the cryostat.

In several magnet designs, especially the high-field LHC dipoles, “quench heaters” are needed which spread out the quench over all magnets in a protection unit to prevent overheating or even burnout of a coil.

Well-built accelerator magnets in the 4-6 Tesla regime achieve the critical field of the superconductor with almost no premature quenches or training. Essential prerequisites for such an optimal performance are a coil and clamp structure providing tight mechanical fixation of each winding and the Rutherford-type cable featuring excellent mechanical stability and a helium-transparent Kapton/glass insulation. Filling the cable voids with solder to increase mechanical stability is not recommendable because it reduces the overall heat capacity. In the 8-9 T regime, the magnetic forces and the stress in the coils are about a factor of two larger than at 5 T and important structural materials like Kapton or the soft copper matrix in the cable are close to their plastic limit. Nevertheless, the LHC dipoles reached the design field of 8.3 T within a few “training” steps.

## References

- [1] H. Brechna, Superconducting Magnet Systems, Springer (1973)
- [2] K.-H. Mess, P. Schmüser, S. Wolff, Superconducting Accelerator Magnets, World Scientific (1996)
- [3] M.N. Wilson, Superconducting Magnets, Clarendon Press (1983), Reissue Oxford U. (1997)
- [4] S. Russenschuck, Field Computation for Accelerator Magnets, Wiley-VCH (2010)
- [5] LHC Design Report, LHC.WEB.CERN.CH/LHC/LHC-DesignReport.html
- [6] A. McInturff et al, PAC 97, p.3214
- [7] S. Sanfilippo et al, LHC Project Report 1091 (2008)
- [8] D. Trbojevic et al, PAC 95, p.2101
- [9] C.P. Bean, PRL 8 (1962) 250; Rev. Mod. Phys. (1964) 31
- [10] H. Brück et al, Z. Physik C44 (1989) 385
- [11] A.K. Ghosh, K.E. Robins, W.B. Sampson, PAC 93, p.2744

- [12] G.H. Morgan, JAP 44 (1973) 3319; A. Devred, T. Ogitsu, KEK preprint 94-156
- [13] A.P. Verweij, H.H.J. ten Kate, IEEE Trans. ASC-5 (1995) 404
- [14] R. Shafer, K.M. Smedley, Proc. Int. Conf. on High Energy Acc. (1992), World Scientific (1993) 298, 629
- [15] M.N. Wilson, Rutherford Lab. internal report RHEL/M1551 (1968)
- [16] N. Schwerg, B. Auchmann, S. Russenschuck, Departmental Report CERN/AT 2007-42 (2007)
- [17] M. Bassetti, C. Biscari in Handbook of Accelerator Physical and Engineering 3rd printing. World Scientific (2006) p.75

## 7.2.5 Superferric Magnets

A. Zeller, NSCL/Michigan State U.

### Advantages over current dominated magnets

1. Iron can be readily machined to great precision and, since the field shape is primarily determined by the iron, very precise fields are possible. Additionally, the profile may be chosen from several possibilities. A sketch of the radial profiles for a window-frame and an H-frame dipole are shown in Fig.1. It is possible to choose a positive or negative sextupole component as shown in Fig.2.

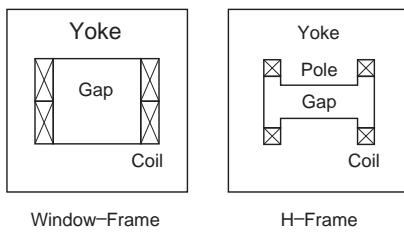


Figure 1: H- and Window frame magnets.

2. Conductor placement is less critical and the field is not sensitive to the macroscopic aspects of the conductor.

3. Current density may be lower because it does not have to be positioned near the magnetic center to be effective. The force on the coil is  $\vec{J} \times \vec{B}$ , which means the force  $\propto$  (current density)<sup>2</sup>. Reducing the force greatly simplifies the coil support structure.

4. For unsaturated iron ( $\mu = \infty$ ) the field  $B$  in the gap  $g$  is

$$B = \mu_0 NI/g \quad (1)$$

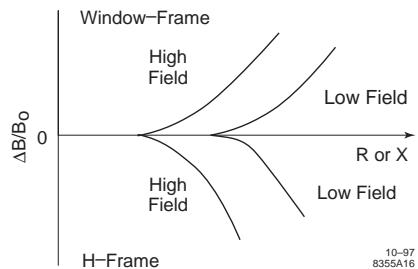


Figure 2: Adjusting sextupole component.

This shows the advantage over resistive magnets in that the brute force method of obtaining a good field region by increasing the gap is obtained by relatively cheap amp-turns.

5. Dynamic range may be extended by using inactive variations such as air gaps, slots, shims and tapered pole tips.

6. Induced supercurrents in the filaments are less of a problem because of the field shaping from the iron.

7. Higher order multipoles may be introduced by shaping the entrance and exit poles of a dipole to give edge focusing or edge curvatures.

8. The iron may be used as all or part of the coil support structure.

9. In contrast to  $\cos n\theta$  type magnets, the effective length is longer than the physical length.

Field limits In principle, there are no field limits except for critical current limits in the conductor. In practice, the limits are set by the field uniformity and dynamic range requirements. If we plot the results of  $B = \mu_0 \mu_{\text{rel}} H$ , where  $\mu_{\text{rel}}$  is the relative permeability of iron (as a function of field) we find three distinct regions, as illustrated below for SAE 1003 iron.

The first region is where  $\mu_{\text{rel}}$  is large. In this region, up to approximately 1.2 T, a properly designed magnet will have the field in the gap increase linearly with current. Highly uniform fields are readily achieved throughout this whole range with a single coil. Similarly, in the high field region above 4 T, the iron acts as if it were a permanent magnet with 2.1 T of field (the saturation magnetization) and the field will again increase linearly with current, but at the air-core slope. It is the middle region, where some of the iron is reaching saturation, that it is difficult to keep the field uniformity high over the whole range. It usually requires multiple coils. One can

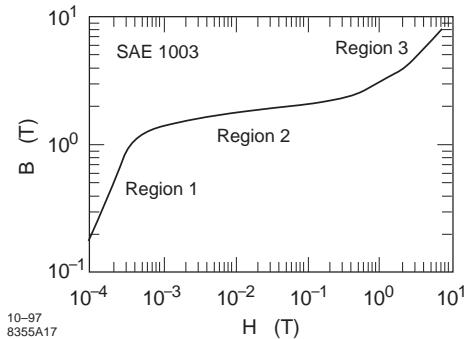
Figure 3: Three regions in  $B$  vs  $H$ .

Table 1 Survey of magnets

| Magnet        | Coil     | Field   | $\Delta B/B\%$     |
|---------------|----------|---------|--------------------|
| NSCL dip.[6]  | single   | 0-1.75T | <0.1               |
| NSCL quad[6]  | single   | 0-35*   | <0.1               |
| TAC dip[7]    | multiple | 0.13-   | ~0.01              |
|               |          | 3.1     |                    |
| g-2 dip[8,9]  | multiple | 1.45    | $5 \times 10^{-5}$ |
| K1200 cyc[10] | multiple | 3-5     | 0.01               |
| Panofsky Q[3] | multiple | 10*     | 0.1†               |

\*T/m; † integral

use the properties of window frame and H-frame dipoles presented in Fig.2 to position the coils.

The above discussion applies to quadrupoles as well as dipoles. Practical limitations, namely the steadily decreasing transfer function, makes a pole tip field limit of 2.5 T in quadrupoles. A hybrid type of quadrupole, the Panofsky quadrupole [1], is a current sheet approximation which derives a significant fraction of the field from the iron. The complications associated with the coil ends and the strict wire and iron placement requirements has limited its use [2, 3]. More details on ferromagnetic materials may be found in [4]. The properties at cryogenic temperatures are given in [5] and in Tab.2.

Tab.1 is a selection of magnets that have been built which illustrate the field range and the uniformity that has been achieved. The uniformity is either the central field or the integral field (or gradients for quadrupoles).

### Dynamic range and time varying fields

Hysteresis effects If we reduce the field from some value in Region 2 in Fig.3 it will not follow the curve back to the origin. This hysteresis effect means the field in the magnet for a given current will vary depending on the history of the magnet. A hysteresis cycle is, however, reproducible and

precise fields can be obtained if the same charging cycle is followed every ramp cycle.

Another potential problem is hysteresis energy loss in the iron. One can integrate the hysteresis cycle to determine this loss,  $W$  (in ergs/cm<sup>3</sup> per cycle) [11] as  $(1/4\pi) \oint H dB$ .

A simplified loss calculation for most magnet grade iron in the annealed state is given by the Steinmetz equation [12, 13]

$$W = \eta B^{1.6}, \quad B \text{ in Tesla} \quad (2)$$

For  $W$  in J/g per cycle,  $\eta = 3.8 \times 10^{-5}$ . Note that  $\eta$  decreases with decreasing impurity content and hardness. Very pure iron, too soft for magnet usage, has a hysteresis loss approximately two orders of magnitude less [4]. For iron at 4 K the hysteresis loss is 1.5 times that at room temperature [14]. Since the concern is for cold iron,  $\eta$  becomes  $5.7 \times 10^{-5}$ . This value is appropriate for most magnet iron (SAE 1001 to 1010) at 4 K.

Hysteresis losses also occur in the conductors and are generally the largest component of loss in modern superconductors. The loss  $W_{\text{cond}}$ , in J/cycle, is [15]

$$W_{\text{cond}} \cong \mu_0 dV J_c \Delta H / 2 \quad (3)$$

where  $d$  is the filament diameter (m),  $V$  the composite volume,  $J_c$  is the average critical current density in the strand at one quarter of the peak magnetic field, and  $\Delta H$  is peak-to-peak external field change in a cycle.

Eddy current effects The energy loss in the iron due to eddy currents is given by  $W_E = cB^2 f^2$ . In general the constant,  $c$ , must be determined empirically. If the magnet has yoke or pole pieces that may be modeled as flat slabs in a parallel magnetic field, then some approximations may allow the calculation of the loss. The skin depth  $\delta$  for iron is

$$\delta = 50.3 (\rho/\mu_{\text{rel}} f)^{1/2} \quad [\text{mm}] \quad (4)$$

where  $\rho$  is in  $\mu\Omega\text{-cm}$ ,  $f$  in Hz. Even though  $\mu$  is a strong function of field,  $\delta$  for any practical field is at most a few cm. If the skin depth is  $\ll$  the slab thickness, the eddy current losses per unit area are approximated by [5]

$$P_E = 0.0634 B^2 \rho / \delta \quad [\text{W/cm}^2] \quad (5)$$

using Tesla,  $\mu\Omega\text{-cm}$ , and mm for units. The total loss is the sum of the hysteresis and eddy current losses.

Another eddy current effect is the slowing down of the decay of the field during a quench.

For magnets built of thin laminations, this is normally a negligible effect (10-20%) [16]. For magnets with solid iron pieces, particularly for iron at 4 K, the effect may be much larger. For a magnet that can be described as roughly circular in cross section, such as the beam line quadrupoles at the NSCL [17], the decay time of the field due to the iron can be approximated as

$$\tau[\text{s}] \approx 4\pi \cdot 10^{-4} \mu_{\text{rel}}/\rho g \quad (6)$$

with  $\rho$  in  $\mu\Omega\text{-cm}$  and the geometrical factor

$$g = \frac{1}{(R_{\text{outer}} - R_{\text{inner}})^2} + \frac{\frac{2\mu_{\text{rel}}}{\pi} \text{gap}}{R_{\text{outer}}^2 - R_{\text{inner}}^2} \quad (7)$$

Note that while  $\mu_{\text{rel}}$  is a function of  $B$ , it is sufficient to use a value which is the average of the permeability at the start of the quench field and approximately half of that field. In any event, the value of  $g$  is only a slowly changing function of  $\mu_{\text{rel}}$ .

Ramp rate effects A major disadvantage of using warm or cold iron in providing the majority of the field for time varying or ac applications is the voltage necessary to produce a constant field ramp rate. The voltage required to produce a change in the current is

$$V = d(LI)/dt \quad (8)$$

In air core magnets  $L$ , the inductance, is constant. In superferric magnets,  $L$  changes with current. At low iron saturation  $L$  is large, as much as a factor of ten greater than the asymptotic  $L$ , therefore the driving voltage changes with time. In particular, the big advantage of superferric magnets is the low current operation - implying large inductances. The high voltages required may produce high electric fields in parts of the magnet not designed for it, such as current leads, if fast field changes are needed. The TAC magnets designed for use in the SSC operated at several thousand Amps, so the inductance was low enough to allow a rapid field change. The K1200 cyclotron, however, has a low field inductance of several hundred Henrys and would require very high voltages to change field rapidly.

**Construction considerations** Materials The only unique material for superferric magnets is the iron. See Tab.2. Keep in mind that iron is brittle at 4 K. This is apparent from the fact that the ultimate yield strength is the same as the ultimate tensile strength. In general, the lower the C content the better the permeability curve. For

Table 2 Iron properties

| Property                                    | 295K | 4K                   |
|---|------|----------------------|
| $\rho$ [ $\mu\Omega\text{-cm}$ ]            | 9.8  | 0.098                |
| $\Delta L/L_{295}$ [mm/mm]                  | 0    | $-202 \cdot 10^{-5}$ |
| $\mu_{\text{rel}}$ max SAE1003 <sup>†</sup> | 5250 | 4900                 |
| Tensile str. [MPa] <sup>*</sup>             | 268  | 744                  |
| † [18]; * [19]                              |      |                      |

dipole and quadrupole pole tips, iron with C content < 0.08% is recommended. This corresponds to SAE 1006 material. It is more important to use low C iron when the amount of iron at 4 K is large in time varying applications because of hysteresis losses. The exact composition of the iron is most important in determining the magnetic properties and should be measured for each application. The other mechanical and electrical properties given above do not depend very strongly upon the C content. Since all energy deposited in the cold mass must be removed by the liquid helium, there is a refrigeration penalty for lossy materials. Low temperature properties of other materials may be found in [5] and Sec.5.1.

Conductors The biggest attraction of superferric is the ability to operate at low current and, hence, smaller helium consumption in the current leads. This requires the use of small conductor. Because the heat transfer from an epoxy impregnated coil package to the LHe is relatively poor, true ac applications are not possible. Therefore, the requirements for very small filament size and Cu-Ni barriers are unnecessary. Additionally, even for dc applications a current density practical limit is  $\sim 20$  kA/cm<sup>2</sup> in a 5 T field. While higher values have been achieved, they are usually not necessary since the purpose of the iron is to avoid having to produce the field only from the coil. True ac applications are possible if one gives up one of the major benefits of a potted coil system: high current density. If we are willing to reduce the current density by a factor of five, cyrostable coils can be built which have excellent heat transfer to the LHe system. Additionally, Cable-in-Conduit coils, used in fusion coils can provide high fields, high current density, excellent heat removal and a solid support structure for the forces on the coil. As yet no superferric CIC magnets have been built. A fast cycling (1 Hz) dipole has been constructed [20] using a technique similar to that used for AC transmission cable. Wires are wrapped around a hollow cupronickel tube and helium flows through the tube to remove the heat (53 W at a ramp rate of 4 T/s).

Tolerances Modern machine shop practices can achieve tolerances on iron pieces of  $< 0.001$  inch ( $25 \mu\text{m}$ ) on very large magnets. Smaller pieces made from solid plate can have tolerances 2-3 times better. Magnets are also constructed from stamped parts, which have small tolerances ( $\sim 10 \mu\text{m}$ ). However, stacking them together tends to introduce tolerance build up and produces  $< 100\%$  packing, especially in large magnets. While this is useful in reducing eddy current loops, it can also produce microscopic imperfections in the field. It is difficult to get wire placement tolerances in current dominated magnets to  $< 50 \mu\text{m}$ , so the intrinsic precision of superferric magnets is a factor of two better. As an example, the S800 spectrograph dipole [6] has a gap volume of  $15 \times 90 \times 350 \text{ cm}^3$  while the variation in the entire gap is  $< 0.025 \text{ mm}$ . Generally, one does not use the volume of field right up to the pole faces. A half gap allowance from the edge of the pole tip will allow the field to “smooth” out so that any remaining machining errors and conductor placement errors are invisible.

Cryogenics Because of the precision to which the cold iron parts may be machined, the placement of the nonmagnetic parts necessary to complete the thermal isolation can result in efficient use of the magnetic field volume. At the NSCL, the superferric quadrupole doublets used in the beam line have an inscribed pole tip radius of 63.5 mm. The warm bore has a radius of 50 mm, while the heat load from the  $\sim 500 \text{ kg}$  cold mass is just sufficient to operate the 20 A leads. Because of the low heat load, and the low current leads, the magnets can be operated in a batch filled mode, reducing the demand on the refrigeration plant.

The biggest demand on the refrigeration system is the cooling required to reduce the iron from room temperature to 4 K. The specific heat of iron at room temperature is  $80 \text{ J/g}$ , so  $\sim 0.05 \text{ kg}$  of helium are required to cool 1 kg of iron from room temperature to 4 K if all of the sensible heat in the helium is used. If only the heat of vaporization of the liquid is used, then  $\sim 4 \text{ kg}$  are required. The actual usage will be somewhere in between those two numbers.

### Operation

Startup The biggest concern in starting up is for magnets which have large amounts of cold iron. Because of the cost in refrigeration, everything must be correct so that costly cool down cycles are avoided. The major problems to be

avoided are mis-wiring the elements, bad coils, short circuits to ground, broken instrumentation wires, thermal shorts and helium leaks. Bad coils can be avoided by careful checking after winding. If the device is a low current, many turn one, then the resistance or inductance checks are not sufficient to reveal all but the largest shorts. Measuring the ac dissipation is the most sensitive method for finding bad coils.

Trouble shooting The most difficult problem to find and fix is usually a helium leak. Once the magnet has been filled with LHe, it is so saturated with helium that the background makes finding all but the largest leaks very difficult. If the magnet is small, or as a last resort leak checking with  $^3\text{He}$  is possible, but expensive.

Failure/replacement Since most superferric magnets will have potted coils, the quench propagation is so rapid that active protection is impossible. Therefore, the magnets are designed to be self protecting. If the calculations are done correctly, the coils will never be destroyed. Furthermore, because there isn't an active quench protection system to fail, once the magnets are operational, there should be no need to replace a damaged one. The most common cause for replacement of a magnet is a helium leak which causes too large of a heat load or raises the coil temperature above the transition temperature. In this case, the best procedure is to scrap the cold vessel and repackage the magnet.

### References

- [1] L.N. Hand, W.K.H. Panofsky, RSI 33 (1959) 927
- [2] K. Tsuchiya et al, NIM 57 (1983) 206
- [3] R. Auzolle, F. Kircher, J. P. Penicaud, PAC 81, p.3228
- [4] R.M. Bozorth, Ferromagnetism, IEEE Press Piscataway (1993)
- [5] R.P. Reed, A.F. Clark, Eds. Materials at Low Temperatures, AMS Metals Park, OH (1983)
- [6] A.F. Zeller et al, Adv. In Cryo. Eng. 37A (1992) 417
- [7] F.R. Huson, IEEE Trans. On Mag. 23 (1987) 1223
- [8] G. Danby et al, NIM A457 (2001) 151
- [9] S.I. Redin et al, NIM A473 (2001) 260
- [10] D. Johnson et al, EPAC 88, p.1261
- [11] E. Warburg, Ann. Physik 13 (1891) 141
- [12] C.P. Steinmetz, Electrician 26 (1891) 261
- [13] C.P. Steinmetz, Trans. Am. Inst. Elec. Engrs. 9 (1892) 3

- [14] J.J. Griesnech, R.L. Powell, *Adv. In Cryo. Eng.* 7 (1962) 303
- [15] H. Brechna, *Superconducting Magnet Systems*, Springer-Verlag (1973)
- [16] R.P. Shutt, BNL Technical Note No.52 (1977)
- [17] A.F. Zeller et al, *IEEE Trans. On Mag.* 27 (1991) 1951
- [18] A.D. McInturff, J. Claus, BNL Report 14630 (1970)
- [19] R.L. Tobler, L.M. Ma, R. P. Reed, *Supercollider III: Proc. 3rd Super Collider Symp.*, Plenum Press (1992)
- [20] A.M. Baldin, et al, *IEEE Trans. On Applied Superconductivity* 5 (1995) 875

### 7.2.6 High Field Accelerator Magnets

*G. Sabbi, LBNL*

High field accelerator magnets are the enabling technology for the highest energy colliders. Progress towards higher fields (Fig.1) is directly correlated to improved conductor properties, the critical current density  $J_c$  being the most important parameter. The ductile Niobium-Titanium alloy (NbTi) has been the superconductor of choice in all projects to date, but its intrinsic properties limit magnet performance to about 10 T at 1.9 K. Among the potential conductors for higher field applications, Niobium-Tin ( $Nb_3Sn$ ) is in the most advanced state of development (see sec. 6.2) [1].  $Nb_3Sn$  wires are available in long lengths and carry currents comparable to NbTi wires of the same size at more than twice the field.  $Nb_3Sn$  is a brittle intermetallic compound with a critical field of 24-25 T at 4.2 K. Because of its brittleness,  $Nb_3Sn$  cannot be drawn into wires like NbTi, but has to be formed in the final geometry by high-temperature heat treatment (650-700 C). In order to use this material effectively, the design concepts and fabrication techniques developed for NbTi magnets need to be modified to minimize conductor strain at all steps. In particular, tight bending of  $Nb_3Sn$  cables at the coil ends, as required for magnetic efficiency, would result in unacceptable critical current degradation. For moderate field applications, alternative design geometries can be considered to increase the bending radii (react-and-wind approach). However, for high field applications coils need to be wound using un-reacted cable, when components are still ductile, and then heat treated to form the  $Nb_3Sn$  phase (wind-and-react approach).

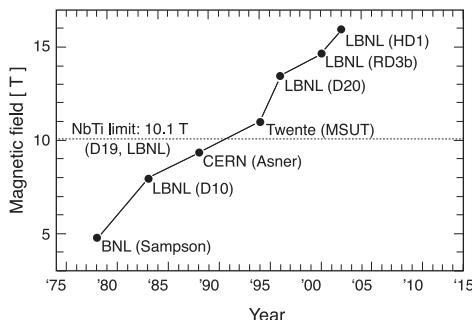


Figure 1: Progress in maximum dipole field with  $Nb_3Sn$ .

#### Magnet design and fabrication

High reaction temperatures require the use of special materials for insulation and coil structural components. With respect to NbTi magnets, polyimide insulation is replaced by fiberglass or ceramics, and metallic parts (bronze, stainless steel, titanium) are used instead than epoxy-glass laminates. In addition, thermal expansion differentials among components could lead to excessive strain on the conductor during the reaction cycle. It is therefore essential to minimize friction between coil and tooling, and to introduce gaps between components at critical locations.

Optimization of  $Nb_3Sn$  Rutherford cables also presents serious challenges. Sufficient compaction is required to maintain mechanical integrity during winding, but may result in deformation of the sub-elements at the cable edges, leading to low-field instabilities and critical current degradation. Acceptable design solutions require lowering the compaction factor especially in the horizontal (wide) direction. Keystone cables require an additional compromise between the vertical compaction factors at the two edges. The resulting range of keystone angles in  $Nb_3Sn$  cables is about half of what would be available in NbTi cables of similar aspect ratio.

Shell-type ( $\cos\theta$ ) coil designs using keystone Rutherford cable have been generally adopted in high-energy colliders due to their self-supporting Roman-arch structure and optimal use of superconductor in the typical design range of interest for aperture and field (see sec. 7.2.4). Using wind-and-react technology this approach has been successfully extended to  $Nb_3Sn$  applications, in particular when large apertures are involved. However as field increases and aperture

decreases, the advantages of shell type coils are progressively diminished. Wide cables are required to minimize the number of layers and magnet inductance. Due to limitations in the keystone angles, coils need to incorporate several wedges, decreasing the magnetic efficiency. Winding issues become critical due limited mechanical stability and high aspect ratios of the cables. High mid-plane stresses due to Lorentz force accumulation can degrade the critical current and limit the magnet performance.

Conversely, block-type coil geometries (Fig.2,3) use flat cables with few or no spacers, resulting in high packing factor and magnetic efficiency even for small aperture designs [2]. When the magnet is energized, physical separation between high-field and high-stress points reduces the sensitivity to stress degradation. Good alignment between cable and field orientation favors efficient grading. Simplification of end part design and coil winding procedures lowers the development time and fabrication cost. These attractive features need to be weighed against the loss of high-field magnetic aperture to provide internal coil support against pre-load. Also, deviations from the simplicity of planar racetrack coils are necessary to address issues of conductor efficiency and field quality. Prototype development is therefore required to evaluate the properties of both  $\cos\theta$  and block-coil dipoles in realistic configurations, and demonstrate performance and cost suitable for accelerator applications.

Large electromagnetic forces are generated in high-field magnets and several new design concepts are being developed to provide adequate coil

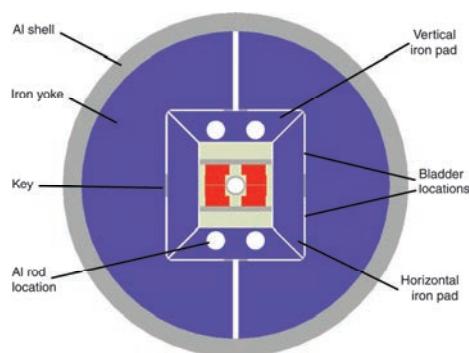


Figure 2: HD2 cross-section.

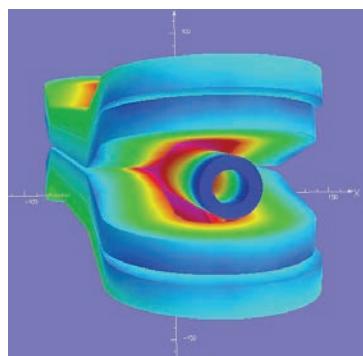


Figure 3: HD2 flared end design.

support. A specific requirement for  $\text{Nb}_3\text{Sn}$  is to minimize stress on the conductor at all stages of magnet fabrication and operation. This can be accomplished using a thick outer cylinder (shell) as the main support element. The shell is typically made of Aluminum and pre-tensioned at assembly using hydraulic elements (bladders) and interference keys. During cool-down the shell tension increases due to differential thermal contraction relative to the iron yoke. This allows to limit the peak coil stress during assembly, and to minimize the amount of structural material taking advantage of its increased strength at lower temperature. To restrain the coil against axial forces, pre-tensioned rods are used to compress thick stainless steel plates against the coil ends. Detailed 3D FEA modeling is required to accurately predict the stress levels at assembly and their evolution during cool-down and excitation (Fig.4) [3].

Protection of  $\text{Nb}_3\text{Sn}$  magnets is complicated by high levels of stored energy, high current density and high critical temperature requiring more heater power and increasing heater delays. In addition, epoxy cracking or damage to the conductor may result from the stress generated during a quench. Extensive heater coverage is required to limit the quench temperatures and voltage within safe limits, and small margins are available to account for failure modes. Nevertheless, prototype tests show that magnets can be adequately protected against quenches.

#### Progress to date

- Dipole magnets Several R&D programs to develop high field magnets using  $\text{Nb}_3\text{Sn}$  have been underway in recent years [4]. The highest dipole fields to date were obtained in the LBNL

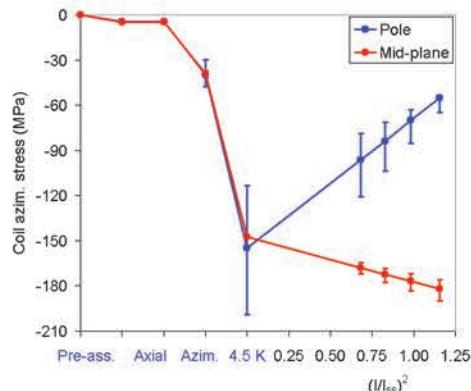


Figure 4: Coil stress evolution in shell-based structures: assembly, cool-down and excitation, for each point range ars indicate the stress variation across the width of the coil.

“HD” magnets using block-coils and a shell-based support structure (Fig.1). The first dipole in this series, HD1, focused on fundamental technology issues using a flat racetrack coil configuration. The magnet achieved a bore field of 16 T at 4.5 K, under coils stresses of 180 MPa. The next step in this series, HD2, incorporates key design features relevant to accelerator applications (Fig.3, 4) [5]. The HD2 short sample dipole field is 15 T at 4.5 K, which can be considered as a practical limit for accelerator quality designs using state of the art Nb<sub>3</sub>Sn conductor.

b. Quadrupole magnets While high field dipoles are critical for increasing beam energies, large aperture focusing quadrupoles are required for high luminosity. The large aperture favors  $\cos 2\theta$  coils for magnetic efficiency and facilitates coil winding, but leads to high stored energy and forces. Minimizing the conductor stress is therefore the primary concern driving all aspects of the design. A series of short models developed by a collaboration of US laboratories approached the conductor limited (short sample) performance under coil stresses approaching 200 MPa. This result confirmed that with optimized coil design and mechanical support, Nb<sub>3</sub>Sn magnets can safely operate under very high stress [6]. A successful scale-up of this design to 4 meter length has been accomplished, providing a key technology milestone to meet the requirements of future colliders such as the high-luminosity LHC [7].

c. High Field solenoids The concept of a muon collider has been proposed as a means to extend the energy reach of lepton machines to the multi-TeV range. The magnet technology requirements of the muon collider ring can be addressed in conjunction with those of high energy hadron colliders. However, before acceleration the muon beam must undergo a “cooling” stage to reduce its emittance by several orders of magnitude. The proposed cooling scheme requires solenoids capable of generating stable fields up to 50 T with reasonably low power consumption [8]. Such field requirements represent a considerable jump with respect to the current state of the art [9]. Several R&D programs are underway to establish the feasibility and limits of this approach [10].

**Future prospects** The most promising candidate superconductors to surpass the intrinsic limits of Nb<sub>3</sub>Sn are the high-temperature superconductors Bi<sub>2</sub>Ca<sub>2</sub>Cu<sub>3</sub>S<sub>2</sub>O<sub>8+x</sub> (bismuth strontium calcium copper oxide or Bi-2212) and YBa<sub>2</sub>Cu<sub>3</sub>O<sub>7</sub> (yttrium barium copper oxide or YBCO). Bi-2212 is available in round wires with isotropic performance and has been made in sufficient lengths for the fabrication of coils based on Rutherford-type cables. Since both Nb<sub>3</sub>Sn and Bi-2212 are brittle and require a wind-and-react approach to coil fabrication, the magnet design methods developed for Nb<sub>3</sub>Sn provide a starting point for work on Bi-2212. However, Bi-2212 technology is significantly more challenging due to higher strain sensitivity, high formation reaction temperature in an oxygen-rich environment, and chemical compatibility of the insulation and construction materials during the reaction heat treatment [11].

YBCO is produced by deposition on thin tapes that include layers of metals, oxides and ceramics for crystal plane alignment, mechanical strength and electrical stability. The main challenge for application of these conductors to accelerator magnets is the development of cables capable of carrying high currents while retaining sufficient flexibility to be wound in coils with minimal degradation. Additional challenges are represented by electromagnetic anisotropy requiring good alignment of the tape to the field direction for optimal current carrying capability, and slow quench propagation velocities requiring new approaches to quench detection and protection [12].

## References

- [1] R. Scanlan in Handbook of Accelerator Physical and Engineering 3rd printing. World Scientific (2006) p.440
- [2] L. Rossi, E. Todesco, IEEE Trans. Appl. Supercond. 19(3) (June 2009) 1186
- [3] P. Ferracin, S. Caspi et al, IEE Trans. Appl. Supercond. 15(2) (June 2005) 1119
- [4] A. Yamamoto, IEEE Trans. Appl. Supercond. 14(2) (June 2004) p. 477
- [5] G. Sabbi et al, IEEE Trans. Appl. Supercond. 15 (2005) 1128
- [6] H. Felice et al, "Test results of TQS03: a LARP shell-based Nb<sub>3</sub>Sn quadrupole using 108/127 conductor", Journal of Physics: Conference Series (2010) 234
- [7] G. Ambrosio et al, "Test results of the first 3.7 m long Nb<sub>3</sub>Sn quadrupole by LARP and future plans", Proceedings of the 2010 Applied Superconductivity Conference
- [8] R.B. Palmer et al, PAC07 p3193
- [9] W. Markiewicz et al, AIP Conf. Proc., Vol. 1218 (2010), pp. 225-230
- [10] R. Gupta et al, "High Field HTS R&D Solenoid for Muon Collider", Proceedings of the 2010 Applied Superconductivity Conference, Washington, August 2010
- [11] A. Godeke et al, Supercond. Sci. Technol. 23 (2010)
- [12] D. Hazelton et al, IEEE Trans. Appl. Supercond. 19(3) (June 2009) 2218

### 7.2.7 Kickers

*C. Burkhardt, SLAC*

Electromagnetic kickers apply elements of the Lorentz force

$$\vec{F} = q(\vec{E} + \vec{v} \times \vec{B}) \quad (1)$$

to impart transverse momentum that deflects the beam from its orbit. The electric field force is dominant at low particle energy, but as particles become relativistic the magnetic field forces become equally effective. The deflection angle,  $\theta$  (rad), imparted on particles of unitary charge with total energy,  $E$  (eV), by an electric field,  $\vec{E}$  (V/m), acting over a length,  $l$  (m), is approximately

$$\theta = \vec{E}l/E \quad (2)$$

and for a magnetic field,  $\vec{B}$  (T), (assuming relativistic particles,  $v \approx c$ ) is approximately

$$\theta = c\vec{B}l/E \quad (3)$$

In general it is more practical to generate large magnetic, rather than electric fields, and so pulsed magnets are commonly used. A notable exception is when very fast response is required, as high gain pulsed magnets usually have an appreciable fill time, and then strip-line kickers are used. Typical uses of kickers include injection/extraction (See Sec.4.12) and orbit correction. The two general classes of kicker are; traveling wave and stationary field. The former is generally used when fast rise and fall times of the deflecting fields are required to avoid unwanted interaction with other particles. The latter provides greater flexibility to increase magnetic gain and achieve large deflection angle.

**Traveling wave systems** Quasi rectangular pulse shapes are generated by travelling wave kicker systems. The most common applications are fast deflection of individual bunches or a train of bunches for injection and extraction in hadron machines and lepton damping rings, several examples are presented in Tab.1. An earlier review is given in [1]. The basic traveling wave kicker circuit is shown in Fig.1. A line-type modulator (Pulse Forming Line, PFL, or Pulse Forming Network, PFN) is charged to a voltage  $U$  and then discharged by an ideal switch via a transmission line and a traveling wave kicker into a termination resistor  $R$ . The kicker can be represented as an additional transmission line whose fields are used for beam deflection. All transmission lines are described by a characteristic impedance,  $Z = \sqrt{L/C}$ , and a propagation time through the line,  $t = \sqrt{LC} = ZC = L/Z$ , where  $L$  and  $C$  are the total inductance and capacitance of the line, respectively. If the lines are loss free and have characteristic impedance,  $Z = R$ , then the discharge pulse is rectangular and reflection free, with voltage  $U/2$  and duration  $T_0 = 2t_{PFL}$ .

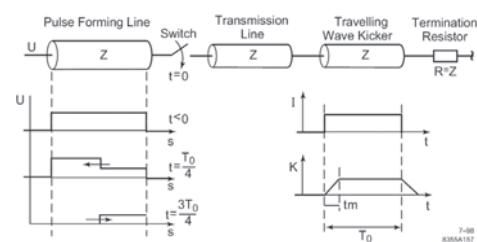


Figure 1: Basic circuit and pulse shapes of traveling wave kicker.

Table 1: Typical parameters of traveling wave kickers (\* strip-line kicker, equivalent magnetic kick strength) (\*\* burst mode repetition time) (\*\*\*) average values, tapered line).

|                           |                          |                           |                         |                 |                 |                 |                |
|---------------------------|--------------------------|---------------------------|-------------------------|-----------------|-----------------|-----------------|----------------|
| Rise time<br>(range)      | ns<br>(%)                | 3.2<br>(0-100)            | 23<br>(10-90)           | 27<br>(10-90)   | 88<br>(5-95)    | 145<br>(5-95)   | 750<br>(0-100) |
| Impedance                 | $\Omega$                 | 50                        | 50                      | 30              | 15              | 12.5            | 5.0            |
| Pulse length              | $\mu\text{s}$            | 0.002                     | 0.075                   | 0.052           | 0.75            | 12              | 6.5            |
| Charge voltage            | kV                       | $\pm 5$                   | $\pm 50$                | 35              | 80              | 60              | 60             |
| Kick strength             | $\text{mT}\cdot\text{m}$ | $1.5 \times 10^{-3}^*$    | 0.15*                   | 3.6             | 88              | 34              | 300            |
| Aperture ( $d \times w$ ) | mm                       | $30 \times 30$            | $50 \times 50^{***}$    | $110 \times 35$ | $251 \times 88$ | $101 \times 61$ | $54 \times 54$ |
| Kicker length ( $l$ )     | m                        | 0.3                       | 9                       | 0.21            | 2.5             | 0.74            | 2.57           |
| Repetition time           | s                        | $3.3 \times 10^{-7}^{**}$ | $2 \times 10^{-7}^{**}$ | 0.01            | 2               | 1               | 15             |
| Reference                 |                          | [3]                       | [4]                     | [5]             | [6]             | [7]             | [8]            |

Traveling wave systems are directional couplers, a result of the cross product in Eq.(1), see for instance [2]. If the kicker fields travel in the same direction as the beam, the magnetic and electric field forces are in opposition. If the kicker is fed from the downstream end, then the forces add, providing the maximum deflection. Hence, a finite time is required for the deflection fields to rise to their equilibrium value. Often called the fill time, this is the time required for the fields to propagate the length of the kicker,  $t_r = t_{KICKER} = \sqrt{(LC)_{KICKER}}$ .

**Strip-line kickers** A straightforward implementation of the traveling wave kicker is a parallel-plate transmission line, often referred to as a strip-line, as illustrated in Fig.2. The fields propagate in free space; hence, the kicker fill time is minimized. A fundamental attribute of the strip-line kicker is that the strength of the electric and magnetic deflection forces is equal, for relativistic particles, so the total kick is given by

$$\theta = 2Ul/Ed \quad (4)$$

for electrode spacing  $d$  (m) (see Fig.2)  $\vec{E} = U/d$  and all other parameters defined as above. Further note that terminating the line into a short (Fig.1,  $R = 0$ ) yields the same kick, since the electric field is zero but the magnetic field is doubled (provisions must be made to remove the reflected energy from the kicker system) [9]. The kick is independent of strip-line impedance, therefore it can be chosen such that the input and output feedthroughs and cables can be standard 50  $\Omega$  devices. Strip-lines kickers generally employ bipolar drive; two PFLs charged to  $\pm U$  that deliver  $+U/2$  to one electrode and  $-U/2$  to the other, both relative to the grounded vacuum chamber ( $\vec{E} = U/d$ ).

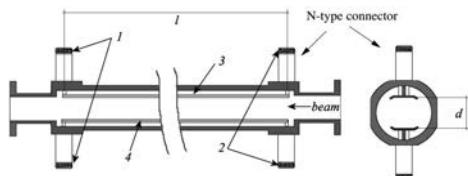


Figure 2: ATF2 strip-line kicker, length  $l = 300$  mm, electrode gap  $d = 30$  mm. (1) Input connectors, (2) Output (load) connectors, (3) +HV electrode, (4) -HV electrode. Courtesy of T. Naito, KEK.

For a given beam energy, the maximum kick that can be generated is set by the applied voltage (typically limited by electrical breakdown along strip-line, in feedthroughs or switch) and kicker length (typically limited by the inter-bunch spacing  $\geq 2t_{KICKER}$ ). If the required kick exceeds the capabilities of a single kicker, multiple units can be used in series. The total kick is the arithmetic sum of the individual units.

Non-relativistic particles, typically low energy ions, are primarily deflected by the electric field as described by Eq.(2). Under these conditions, the kicker structure is no longer directional so the kicker can be fed from the upstream end and the fill time is approximately zero. If the phase velocity of the kicker fields can be matched to the particle velocity, the kicker length is no longer limited by the inter-bunch spacing. Typically, strip-line electrodes with increased path length (e.g. meander [10]) are used to reduce the phase velocity.

**Pulsed magnets** High permeability materials in the flux return path (refer to Figs.3 and 4) can be used to increase the magnetic field, and associated kick. Adding a flux return core to a

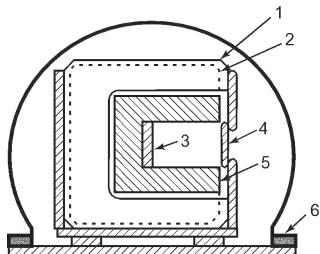


Figure 3: Schematic cross-section of C kicker with parallel plate matching capacitors in vacuo. (1) Capacitor plate (ground potential), (2) Capacitor plate (high voltage potential), (3) Current conductor (high voltage potential), (4) Current conductor (ground potential), (5) Magnet yoke, (6) Vacuum enclosure.

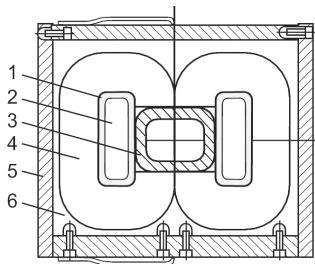


Figure 4: Schematic cross-section of window frame kicker with wound tape core and ceramic chamber. (1) Insulation, high voltage coil, (2) Conductor, high voltage coil, (3) Ceramic beam pipe, (4) Magnetic core, (5) Mechanical frame, (6) Epoxy reinforcement.

vacuum-field strip-line of impedance  $Z$  increases the magnetic field by approximately,  $\mu_0 c / Z$ , where  $\mu_0 c = 377 \Omega$  (impedance of free space). As noted previously,  $Z$  is typically  $\sim 50 \Omega$ , so the magnetic gain can be substantial.

The introduction of the high permeability material increases the inductance of the kicker. To maintain similar impedance, a corresponding increase in capacitance is required. Ideally, both the inductance and capacitance would be uniformly distributed. However, real magnets are in general assembled as lumped ladder networks. They are normally built up of  $\sim 5\text{-}30$  cells, depending on pulse response requirements. With a magnet inductance  $L_m$ , and  $n_m$  cells per magnet, the cell inductance,  $L_c = L_m/n_m$  and the corresponding cell capacitance, required to achieve impedance  $Z_o$ , is  $C_c = L_c/Z_o^2$ . However, because of the dispersion introduced by the lumped element

network, the characteristic impedance is frequency dependent, acting as a low pass filter. The reader is referred to the first edition of this handbook [11], for detailed response calculations.

Travelling wave kicker magnets are constructed in "window frame" or "C" configuration. In the window frame approach both conductors are inductive and the equivalent circuit is a four terminal network. In a C configuration (Fig.3), the return conductor is not surrounded by flux thus reducing the four terminal to a three terminal network, facilitating the use of ground referenced coaxial connectors at the input and output terminals. The entire magnet is housed in vacuum. For beam impedance reasons, it is often preferred in modern synchrotrons to equip kicker magnets with thin metalized ceramic vacuum chambers placed into the magnet aperture rather than placing the entire magnet into a vacuum enclosure. (See [12] and Secs.5.10.8 and 6.15.2) A cross section of such a window-frame type magnet with a tape wound steel yoke is shown in Fig.4.

**Line-type modulators** For a pulse duration  $\leq 3 \mu\text{s}$ , usually coaxial high voltage lines, length  $l_c$ , are employed as PFLs. Under matched conditions,  $Z_{PFL} = R_{KICKER}$ , the pulse duration is twice the propagation time through the line with permittivity  $\epsilon_r$

$$T_o = 2l_c \sqrt{\epsilon_r/c} \quad (5)$$

In general, the pulse rise time will be limited by parasitic series inductance in the system,  $L_s$ , to

$$t_g = 3L_s/2Z \quad (0\text{-}95\%). \quad (6)$$

In a well designed system, the stray inductance can be reduced to approximately that of the switch. For coaxially mounted three gap ceramic thyratrons  $L_s \approx 100\text{-}150 \text{ nH}$ . The fall time of the pulse is slightly longer, depending on cable length and dispersion.

For a pulse duration  $\geq 3 \mu\text{s}$ , the flat-top droop caused by losses to the series resistance of the line becomes significant and a PFN, the lumped element approximation of a transmission line, is used. A circuit diagram of a complete PFN-powered kicker system is given in Fig.5. Following from the PFL discussion, the pulse duration into a matched load is twice the propagation time, which is

$$T_o = 2n_g \sqrt{L_g C_g} \quad (7)$$

for a PFN with  $n_g$  stages and stage inductance,  $L_g$  and capacitance,  $C_g$ , and the PFN impedance is

$$Z = \sqrt{L_g/C_g}. \quad (8)$$

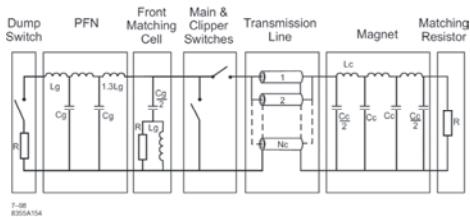


Figure 5: Schematic diagram of traveling wave magnet kicker system ( $R = Z$ ).

As the number of stages is increased, the magnitude of the flat-top ripple and the ratio of pulse rise time to duration is decreased. Typically, the number of stages chosen varies between about 10 and 30. If the natural pulse rise time of a PFN is too long, a speed-up or “front matching cell” is added. Due to dispersion, the natural fall time of a PFN is often a large percentage of the pulse duration. If a short fall time is required, the output of the PFN is short-circuited by means of a second “clipper” switch and a third “dump” switch or a diode assembly, in series with a matched resistor, is connected to the open end of the PFN to dissipate the residual energy in the PFN to avoid oscillations. The relative timing between the three switches controls the duration of the output pulse. Alternative PFN derivations are discussed in [13].

The stage inductors are either wound as individual one-layer solenoids or as a continuous one-layer coil over the whole PFN length. In the former design, magnetic coupling between cells is, to a first approximation, negligible and the pulse shape can be adjusted by varying the inter turn distance of the solenoids. The latter design provides magnetic coupling between stage inductances that is equivalent to a negative inductance in series with the capacitor, which can offset its internal inductance. However pulse shape adjustment is not practical with the continuous coil approach, so the design and its implementation must be very accurate. The design of low impedance, low flat top ripple PFN’s is treated in [8].

Typically CAD tools are used to achieve the design accuracy required for kicker and modulator applications. Circuit analysis tools, Ref.[14] provides a thorough discussion of available options, are extremely useful for determining component values. Care must be taken to accurately model the components (e.g. capacitor ESL and ESR and inductor inter turn capacitance).

Additionally, numerical tools for electrostatic and electromagnetic field calculations, see Ref.[15], are helpful for magnet designs and determination of high voltage insulation requirements.

**Stationary field systems** Fast deflection is most economically accomplished by a capacitive discharge into a purely inductive magnet. Compared to a travelling wave kicker system, the PFN reduces to a single capacitor,  $C$ , and the magnet to a single inductor,  $L$ , as illustrated in Fig.6. If, as is generally true, resistive losses can be neglected, the current waveform is sinusoidal. The peak current,  $I_{peak} = U\sqrt{C/L}$ , is reached at time,  $t_r = \pi/2\sqrt{LC}$ . Compared to the matched travelling wave system, the same deflection strength is achieved at half the charging voltage. There are three common variants of the basic  $LC$  circuit which differ in temporal behavior after the deflecting field maximum; half-sine, full-sine, and exponential decay, are described briefly below. A more detailed discussion is provided in [11].

Half-sine kicker These are generally used for the deflection of a single bunch, at the peak of the magnet current. Most common applications are injection and extraction in accelerators with only a limited number of bunches or in lepton machines with accumulation schemes for injection [16]. The half-sine pulse is achieved with the basic circuit in Fig.6 (none of the grayed-out components);  $L$ ,  $C$  and asymmetric switch,  $S$ , with associated transient suppression snubber,  $R_S C_S$ . At the current zero crossing, at time  $2t_r$ , the switch commutes and terminates the pulse.

Full-sine kicker In applications where reversal of the deflecting fields is acceptable, such as low duty-cycle linear machines and septum magnets, the full-sine kicker offers advantages over the half-sine topology. As illustrated in Fig.6, a reverse diode,  $D_S$ , is added to allow the magnet current to return to the capacitor,  $C$ . During current reversal the thyristor commutes, so the

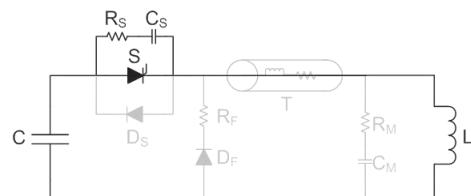


Figure 6: Simplified schematic diagram of LC kicker system.

pulse ends after one sinusoidal period. Average power consumption is reduced, as compared to the half-sine approach, as the magnet energy (less losses) is recovered in the energy storage capacitor,  $C$ .

Exponential decay Fast rising pulse shapes with long exponential decay are used in fast extraction systems with relaxed flat-top requirements. The main application is beam abort (dumping) systems [17]. A freewheeling diode,  $D_F$ , is added to the basic LC circuit in parallel to the magnet load, shown in Fig.6. During the sinusoidal current rise  $D_F$  is reverse biased and does not conduct. At maximum current  $D_F$  becomes forward biased,  $S$  is reverse biased and the magnet current returns through  $D_F$  instead of  $C$ . The magnet current decays with a time constant,  $t_F = L/R_F$ .

In general, the stationary field system modulator is placed in the accelerator tunnel adjacent to the magnet as any interconnecting line will degrade the magnetic pulse, in magnitude and/or harmonic content. In applications where the modulator cannot be directly mounted to the magnet, the interconnecting cable can distort the waveform. If the transit time through the cable,  $t_t$ , is short compared to the pulse rise time,  $t_r \geq 10t_t$ , the cable inductance adds to the load (as illustrated in Fig.6), diminishing the peak current and magnetic field. If the rise time is faster, impedance mismatches at the cable ends will cause reflections that will introduce perturbations in the magnet current. Such reflections can be minimized by a terminating filter at the load with  $R_M$  equal the cable impedance and  $R_M C_M > t_r$ .

Magnet Stationary field magnets can be constructed similarly to traveling wave devices. However, multiple turns can be employed to increase the magnetic gain of the circuit. With the potential for such a large magnetic deflection, often these kickers do not incorporate any electrostatic deflection. Multi-turn coils can also be used in lieu of magnetic materials. These air-core magnets are employed in applications where the residual field from magnetic materials cannot be tolerated.

### Components and materials

Switches (See also Sec.6.18.1) Both gas and solid state switches are in common usage for pulsed magnet applications. Generally, much greater power (both higher voltage and current)

can be switched with gaseous devices. However, topologies to array multiple solid state devices have been demonstrated across the parametric space of most pulse magnet applications. Relatively speaking, gaseous switches are a technology in decline relative to solid state, with ever decreasing device availability and increasing costs. Further, as switch reliability becomes increasingly important to meet system availability needs for user facilities and terra-scale accelerator systems, the longer life of solid state devices is becoming increasingly important.

Gas switches Although pressurized spark gaps and vacuum tubes find some limited use in pulse magnet systems, thyratrons are the most common gas switch used in critical, high performance applications. They offer reliable switching of high voltage (to  $> 100$  kV), at high current (typically  $\sim$  few kA), at high repetition rate (to  $> 10$  kHz), with low turn-on jitter ( $< 10$  ns typical), with a large dynamic voltage operation range (typically down to  $\sim 0.5$  kV), and are switched with a low power trigger ( $\sim$  kV,  $\ll J$ ). Thyratrons operate as a closing switch, they will not interrupt current flow. Switching is rapid (as fast as a few ns, although  $\sim 0.1$   $\mu$ s is more typical) and will support a large dI/dt ( $> 10$  kA/ $\mu$ s is typical). Voltage reversal of greater than a few kV should be avoided (there are specialized tubes designed to support significant voltage reversal) as it can induce internal arcing and grid damage. Thyratrons require external electrical supplies for the cathode heater and the reservoir (gas pressure regulation), typically  $\sim 0.1$  kW each, but can usually be air cooled.

Solid state switches Four classes of semiconductor device are finding use over the broad parametric range of pulse magnet switches; thyristor, insulated gate bi-polar transistor (IGBT), metal oxide semiconductor field effect transistor (MOSFET), and diode (used as an opening switch). Thyristors operate as closing switches to 6.5 kV (SiC devices hold the promise of extending this to 20 kV) and  $> 10$  kA. The fastest of these devices (evolved from the GTO, but marketed under varying designations including IGCT and pulse power optimized thyristor) switch in  $\sim 1$   $\mu$ s, support dI/dt  $> 10$  kA/ $\mu$ s, and can be force commutated (although often only when the cathode current is small). The switching time of IGBT power modules is approximately 0.5  $\mu$ s ( $\sim 0.1$   $\mu$ s for individually packaged dies), but dI/dt is

limited to  $< 10 \text{ kA}/\mu\text{s}$  and commutation current to a few kA to avoid destructive oscillations in the module package. The power MOSFET has an intrinsic switching time of  $\sim \text{ns}$ , but due to parasitic packaging elements 10 ns is more typically achieved. However, few devices are available at voltages higher than 1.2 kV and peak currents are typically no more than  $\sim 0.1 \text{ kA}$ . Therefore, large arrays of devices are required for all but the lowest power pulse magnet applications. The alternative for ns switching is to use a diode opening switch in a current-charged PFL topology, see [18]. The packaging technology to stack dies to form a high voltage diode is well developed and modulators to  $> 20 \text{ kV}$  into 50 ohm have been demonstrated using this topology [19]. The major challenge is to maintain balanced energy transfer that minimizes pre and post pulses at the pulse magnet.

Ferrite Ni-Zn ferrite is the standard yoke material for kicker magnets. The magnetic permeability is sufficiently high to neglect in first approximation the reluctance of the yoke with respect to that of the beam aperture. The electric conductivity is low enough to allow direct contact with the current conductors avoiding the need of conductor insulation. Properly cleaned ferrite has a low outgassing rate, acceptable for UHV enclosures. The CMD5005 material is typical of ferrite for this application; 2100 initial (relative) permeability, 5500 maximum permeability, 3.3 kG saturation flux density, 0.12 Oe coercive force, and  $10^8 \Omega\text{-m}$  bulk resistivity. Ferrite properties are typically evaluated at 10 kHz, therefore under fast kicker conditions (particularly during the magnet fill time) the permeability may be substantially less, which will affect the driver matching.

Laminate cores For magnetic field pulses with an induction  $> 250 \text{ mT}$  thin steel laminations are used as yoke material. For an acceptable pulse response the lamination thickness,  $d$ , is chosen to less than the skin depth (See Sec.7.2.3.3). As cost decreases with increasing  $d$ , material with high resistivity and relatively low permeability is preferred. Traditionally, 3% Si-Fe alloys have been used when thicker laminates are acceptable and Ni-Fe alloys when  $d \sim 25 \mu\text{m}$  is required (Si-Fe is typically not available thinner than 50  $\mu\text{m}$ ), both have bulk resistivity around 50  $\mu\Omega\text{-cm}$ . Although more costly, amorphous and nanocrystalline material bulk resistivity is  $> 100 \mu\Omega\text{-cm}$ , and these are typically cast in 25  $\mu\text{m}$  thickness. For  $d < 100 \mu\text{m}$  the laminations are stamped with

the shape of the yoke profile and stacked successively to form the magnet. Fast applications employ stacks of wound tape cores with a tape thickness down to 25  $\mu\text{m}$  into which the beam aperture is cut [20].

## References

- [1] E.B. Forsyth, M. Fruitman, PA 1 (1970) 27
- [2] D.A. Goldberg, G.R. Lambertson, AIP Conf. Proc. 249 (1992) 537
- [3] T. Naito et al, NIM A 571 (2007) 599
- [4] E.G. Cook and P.L. Walstrom, PAC 03, p.544
- [5] T. Fowler et al, PAC 89, p.408
- [6] D. Fiander et al, PAC 81, p.2949
- [7] H. Kuhn, G.H. Schröder, Proc. 14th IEEE Pulse Power Modulator Symp. (1980) p.264
- [8] L. Ducimetiere et al, Proc. 10th IEEE Pulsed Power Conf. (1995) p.1406
- [9] C. Pappas, Proc. 25th IEEE Power Modulator Symp. (2002) p.110
- [10] F. Caspers et al, Proc. EPAC02, p.873
- [11] G.H. Schröder in Handbook of Accelerator Physics and Engineering, 1st ed 3rd print World Scientific (2006) p504
- [12] M. Mayer et al, EPAC88, p.1378
- [13] G.N. Glasoe, J.V. Lebacqz, Pulse Generators, McGraw-Hill (1948)
- [14] F. Jenni, Simulation Tools, CERN Accelerator School (2004)  
<http://cdsweb.cern.ch/record/987562/files/p363.pdf>
- [15] U. van Rienen et al, Scientific Computing in Electrical Engineering, Springer-Verlag (2000)
- [16] G.H. Schröder et al, 16th IEEE Pulsed Power Conf. (1984) p.103
- [17] E. Carlier et al, Proc. 9th IEEE Pulsed Power Conf. (1993) p.865
- [18] F. Arntz et al, PAC07 p2972
- [19] D. Wildman et al, PAC 01 p.3726
- [20] M. Mayer et al, IEEE Electrical Insulation Magazine 12(6) (1996) 15

### 7.2.8 Permanent Magnet Elements

*K. Halbach (deceased), R. Schlueter,  
LBNL*

These devices can be classified by configuration and material as: (i) iron-free devices using only current (or charge) sheet equivalent material (CSEM), i.e. permanent magnet (PM) material; (ii) hybrid devices using CSEM and iron or equivalent high permeability material; and (iii) CSEM assisted electro-magnets. They can also

be classified by device function as: (a) focusing and optics correction magnets (multipoles); (b) bend (dipole) magnets; (c) insertion devices; and (d) special purpose magnets (e.g. to produce solenoidal fields). (See Sec.6.4 for PM material properties.)

The advantages of PM systems with respect to electro-magnet (EM) systems are: (i) stronger fields than EM systems. (Current density in EM systems scales like inverse linear dimension if one keeps the field constant, leading at some critical dimension to unmanageable cooling problems; PM systems do not suffer from this limitation, leading, below a critical dimension, to stronger fields. This applies to all PM systems.); (ii) Iron free PM's are immersible in other fields without perturbing them. (iii) Most PM systems are compact and light. (iv) Absence of power supplies or high capacity cooling systems lead to greater reliability and convenience. (v) CSEM can be classified as "analytical material". Because the properties of CSEM are very simple, design of PM systems is fairly simple (see last section).

The principal disadvantage of PM systems is that the fields cannot be changed or can be changed only slowly.

### Iron free multipoles

Ideal system properties If one were to place CSEM (with a remnant field of  $B_r$ ) between two infinitely long concentric cylinders with radii  $r_1$  and  $r_2$ , and with the easy axis at location  $(r, \phi)$  pointing in the direction  $(N+1)\phi$ , complex potential  $F = A + iV$  ( $A$  and  $V$  representing vector and scalar potentials in appropriately defined units) and field would be given by, with  $z = x + iy$  and  $B^* = B_x - iB_y$ ,

$$F = -iB_r r_1 \left(\frac{z}{r_1}\right)^N \frac{1 - (r_1/r_2)^{N-1}}{N-1} \quad (1)$$

$$B^* = B_r \left(\frac{z}{r_1}\right)^{N-1} \frac{N[1 - (r_1/r_2)^{N-1}]}{N-1} \quad (2)$$

When one has this kind of a system, but CSEM only between two planes oriented perpendicularly to the system axis, separated by the distance  $L$ ,  $F$  and  $B^*$  integrated from  $-\infty$  to  $\infty$  in the axial direction would be given by the above expressions, multiplied by the physical length  $L$ . Fields in the immediate vicinity of the axis of the system as a function of location along the axis can be expressed analytically and have been published for quadrupoles [1].

Properties of imperfect systems Producing CSEM in the form needed for the system described above with sufficient quality is presently not possible. However, blocks of CSEM with uniform remanent field (both regarding strength and direction) can be produced with high precision, leading to the design of the segmented multipole. This structure is essentially the same as described above but broken up by radial lines into  $M$  geometrically identical blocks, with the uniform  $\vec{B}_r$  oriented as the center line of each block should be oriented. Eq.(2) then becomes [2]

$$B^*(z) = B_r \sum_{v=0}^{\infty} \left(\frac{z}{r_1}\right)^{n-1} \frac{n}{n-1} \left[1 - \left(\frac{r_1}{r_2}\right)^{n-1}\right] K_n \quad (3)$$

where  $n = N + vM$ ,  $N$  being the fundamental multipole order, and  $K_n$  a factor describing the consequences of the segmentation.  $B_r$  is a complex number, with the phase freely choosable. For the just described system, but with the size of each angular slice being  $2\pi\epsilon/M$  ( $\epsilon \leq 1$ ),

$$K_n = \sin[(n+1)\epsilon\pi/M]/[(n+1)\pi/M] \quad (4)$$

If the segments have trapezoidal cross-section, with  $r_1$  and  $r_2$  describing the shortest distance from the magnet center to the inner and outer trapezoidal boundaries, then

$$K_n = \cos^n(\epsilon\pi/M) \sin(n\epsilon\pi/M)/(n\pi/M) \quad (5)$$

Fig.1 shows such a quadrupole with  $M = 16$ .

From these equations (and experience) we conclude the following: (i) Provided the CSEM is not demagnetized, these magnets provide remarkably strong fields. Example: a quadrupole with  $M = 16$ ,  $r_2/r_1 = 3$  and  $B_r = 1.1$  T will, with Eq.(5), have 1.84 T at the aperture; (ii) The first undesired harmonic is of order  $N + M$  and is rather strong at the aperture radius. But because of the high order of this harmonic, it becomes weak quickly when going toward the center of the quadrupole. In addition, by choosing  $\epsilon$  well, this harmonic can be made zero; (iii) Due to the slightly larger than 1 (typically 1.03-1.1) differential permeability of the CSEM, there will also be harmonics present beyond the ones listed above. The lowest order of such harmonics is usually determined by the degree of invariance (ignoring signs) of the geometry of the magnet under rotation about the axis, usually three times the order of the fundamental. These harmonics are typically <1% at the aperture. These and length-integrated harmonics due to other imperfections

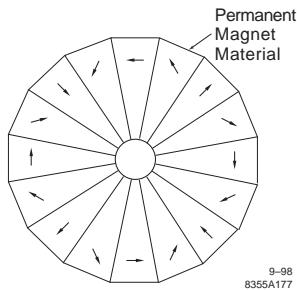


Figure 1: CSEM quadrupole,  $M = 16$  with trapezoidal blocks of CSEM (see also Sec.5.5).

can be canceled by a system of correctors [3], which consists of a number of pairs (to give control over both correction amplitude and phase generated by the corrector) of round, rotatable blocks of CSEM disposed around the perimeter of the magnet.

### Iron-free insertion devices and other structures

Conventional insertion device In the limiting case of  $N, r_1, r_2 \rightarrow \infty$  (keeping  $r_2 - r_1$  constant) we obtain an infinite linear array. Putting two of these together forms an iron-free insertion device. Fig.2 shows a structure with  $M' = 4$  blocks per period. The transverse field  $B^*$  as a function of distance into the paper plane is

$$B^* = i2B_r \sum_{v=0}^{\infty} \cos(nkz) e^{-nkh} \times \left(1 - e^{-nkL}\right) \frac{\sin\left(\frac{n\pi}{M'}\right)}{n\pi/M'} \quad (6)$$

$$n = 1 + vM'; \quad k = 2\pi/\lambda$$

Utilizing PM magnetization in the third dimension enables iron-free device strengths even greater than those given by this two-dimensional formula [4].

Other iron-free devices Since the introduction of the conventional iron-free insertion devices, many periodic insertion device structures with special properties have been built and used with great success. Foremost among these is the permanent magnet elliptically polarized undulator [5] which enables a periodic field of arbitrary polarization (linear, helical, circular) and orientation [5]. Note that this technology is also suited to build other periodic (like periodic solenoidal fields [1] magnets) and nonperiodic structures

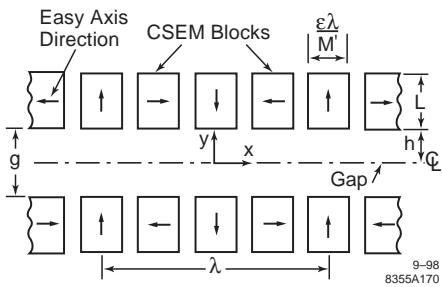


Figure 2: Iron free insertion device.

(like septum magnets with an “immortal” septum that has instead of a material septum a vacuum region with very distorted fields), but their use is also infrequent.

### Hybrid (CSEM + iron) devices

Generic qualitative design philosophy The driving thought behind hybrid magnets is the desire to have a magnet where the field distribution is controlled by high  $\mu$  material (making  $\mu$  irrelevant as long as  $\mu \gg 1$ ) and combining that with the other advantages of PM systems, like high field levels for small magnets, reliability, etc. One of the most expedient design/construction philosophies consists of attaching to the field shaping iron configuration CSEM in such a way that the iron poles are put onto the scalar potentials needed to produce the desired fields, preferably without generating the direct fields discussed below. Under most circumstances it is advisable to enclose that structure by a yoke that shields the inside against outside fields, and the outside against fields “leaking” from the inside. Fig.3 shows a hybrid dipole example.

Quantitative design procedure Task definition: The analysis/design of a hybrid magnet can be formulated as follows: One has to satisfy  $\nabla \times \vec{H} = \nabla \cdot \vec{B} = 0$  subject to the materials equation

$$\vec{B}(\vec{H}) = \vec{B}_r + \mu \vec{H} \quad (7)$$

for CSEM, and the boundary conditions at the interfaces between high  $\mu$  material and CSEM and/or vacuum. For convenience,  $\mu_0$  is included in  $H$ , i.e.  $H$  is measured in the same units as  $B$ , namely T or G. The passive part ( $\mu \vec{H}$ ) of the  $B(H)$  relationship for the CSEM is, in reality, a tensor relation because of the slightly different values of  $\mu$  parallel and perpendicular to the easy axis. Since these differential permeabilities are so

close to one, we ignore that distinction and use the implied scalar relationship. We assume further that the fields are perpendicular to all high permeability bodies, i.e. we use  $\mu = \infty$  for these bodies. This is a very good approximation whenever  $\mu \gg 1$ , a condition one wants to satisfy if at all possible to make the device independent of the magnetic properties of the high  $\mu$  materials. Finite permeability effects are usually small enough to be ignorable, and can be taken into account at the end of the design.

Re-formulation of problem: Using  $\vec{H} = -\nabla V$  to describe all fields and related quantities, inserting the materials equation into  $\nabla \cdot \vec{B} = 0$  gives

$$\nabla \cdot \mu \vec{H} = \rho_m = -\nabla \cdot \vec{B}_r \quad (8)$$

This means that instead of dealing with a material described by an extremely anisotropic materials equation, we are dealing with a “normal” isotropic material, and magnetic space charges that are sources of magnetic fields. This helps both visualizing what the material does, and makes the quantitative description remarkably simple. The visualization is enhanced by the fact that we are practically always dealing with material with a constant  $\vec{B}_r$ , yielding magnetic charges only on the CSEM body surfaces. Since it is not at all unusual to have to design 3-D structures, the ability to derive fields from a scalar potential (instead of, for instance, a vector potential) greatly simplifies the analysis/design.

Direct and indirect fields and related quantities: Understanding the nature and properties of fields in a hybrid magnet is greatly simplified by distinguishing clearly between direct and indirect fields, defined as follows:

(i) Direct fields (and related quantities) are those fields caused by the magnetic charge terms in Eq.(8) when all high  $\mu$  surfaces are on zero scalar potential.

(ii) Indirect fields are those fields caused by high  $\mu$  surfaces being on non-zero scalar potentials (relative to one arbitrarily chosen surface that is always on zero scalar potential), with all charge terms ignored. Implied in these definitions is that the CSEM permeability has to be considered in the calculation of both types of field, an easy task because  $\mu$  is rarely larger than 1.1.

The flux balance equation: Clearly, the total field everywhere can be obtained by linear superposition of all direct and indirect fields. Integrating over the surface of each  $\mu = \infty$  surface

(except the surface chosen as the zero potential reference surface) the flux density associated with the direct fields caused by all of the CSEM will give the direct flux for each of these  $\mu = \infty$  bodies, and it is convenient to arrange these flux values in form of a direct flux array (or vector)  $\Psi_d$ . Even though under many circumstances these flux determinations are trivial, general formulae for calculating them will be given below. Integrating similarly the flux going into each body caused by the indirect flux generated by putting these bodies on arbitrary potentials  $V_{in}$  will lead to a similar flux array  $\Psi_i$  that is related to the  $V_i$  array by  $\Psi_i = CV_i$  with  $C$  representing a symmetrical capacity matrix that is in every respect equivalent to such matrices used for electrostatic calculations. The symmetry of the  $C$  matrix makes it legitimate to represent the magnetic interaction between iron surfaces on differing potentials by analog capacity networks, again just as one does for electrostatic systems. Having calculated  $\Psi_d$  and demanding that the total flux (equaling the sum of the direct and indirect flux) into each of the  $\mu = \infty$  bodies is exactly zero gives

$$CV_i + \Psi_d = 0 \quad (9)$$

allowing the calculation of  $V_i$ , and with it the fields everywhere. For many systems (see below) the matrix  $C$  has only one element. When more than one capacity is needed, it is informative to represent these capacities, together with the direct flux induced by the charges, by equivalent analog circuits. Note that when one deals with 3-D structures, the components of  $\Psi_d$  have the dimensions  $T \cdot m^2$  and the components of  $C$  have the dimensions  $m$ . When one is dealing with 2-D structures (a good starting point to get the feel of things),  $\Psi_d$  components have the dimensions  $T \cdot m^2/m = T \cdot m$ , resulting in dimensionless capacities.

Direct flux calculation: The direct flux induced by CSEM onto a  $\mu = \infty$  body, or any part of such a body, can be calculated as follows: if  $V_i$  and  $H_i$  are the indirect scalar potential and field associated when putting the surface of interest onto potential  $V_0$  and all other surfaces onto zero scalar potential, the direct flux “deposited” onto that surface is given by following CSEM volume integrals

$$\Psi_d = - \int \vec{B}_r \cdot \frac{\vec{H}}{V_0} dv = \int \rho_m \frac{V_i}{V_0} dv \quad (10)$$

The second form gives the not surprising answer that when uniformly magnetized CSEM touches a

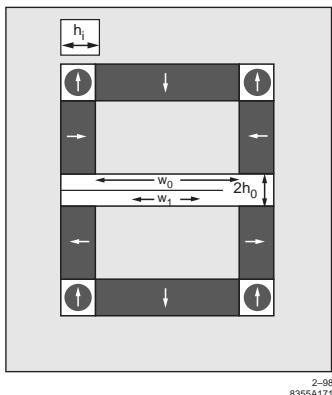


Figure 3: Hybrid dipole.

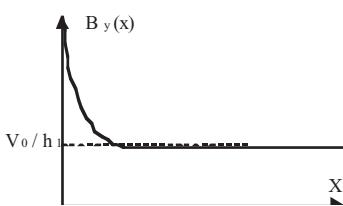


Figure 4: Field near pole corner.

surface, the whole flux associated with the surface charge is deposited onto that surface (with the direct field not visible!), making the direct flux calculation under many circumstances very simple.

Indirect flux (capacity) calculation: Fig.3 depicts a hybrid dipole magnet, the 2-D flux (i.e. flux per unit length in the direction into the paper) entering the pole over a uniform field width region  $w_1$  on scalar potential  $V_0$  from the median plane at distance  $h_0$  on  $V = 0$  is  $\Psi_i = V_0(w_1/h_0)$ .

Toward the corners of the pole, the field clearly increases, as shown schematically in Fig.4 for the region close to the left corner of the pole. Fig.5 shows the corresponding flux as a function of distance  $x$  from the edge of the pole.

Since the offset of the asymptote to the flux curve from the uniform field part of that curve  $\propto V_0$ , the flux entering the pole between the left edge and a location  $x$  to the right of the edge can be written as  $\Psi_i = V_0 [(x/h_0) + E]$  with  $E$  a dimensionless number dependent only on the geometry. Because of the exponential decay of non-uniform fields caused by the corner,  $x$  does not have to be very large for this formula to hold.

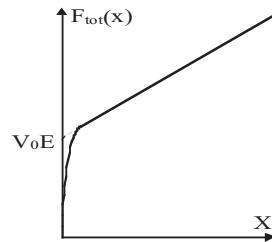


Figure 5: Flux on pole face near corner.

The excess flux coefficient  $E$  can be obtained in many ways, with analytical formulae being the most convenient to use. From this follows for the contribution of the pole face of width  $w_0$  to the total capacity,

$$C = w_0/h_0 + 2E_{PF} \quad (11)$$

$E_{PF}$  for each corner of the pole face is ( $b = h_1/h_0$ ) given by

$$E_{PF} = \frac{1}{\pi} \left\{ \ln \left[ 0.25 \left( 1 + \frac{1}{b^2} \right) \right] + 2b \tan^{-1} \frac{1}{b} \right\} \quad (12)$$

The excess flux into the other side of a corner is given by the same formula, but with  $b$  replaced by  $1/b$ . The flux into the midplane (excluding the flux through the iron yoke in the midplane) is

$$\Psi_{MP} = V_0 [(w_0/h_0) + 2E_{MP}] \quad (13)$$

with  $E_{MP} = E_{PF} + 2 \ln(b)/\pi$ .

Analysis/design equations for a box dipole: For a dipole magnet of the type shown in Fig.3, the direct flux is

$$\Psi_d = B_r \sum \{ \text{area of CSEM touching the pole} \} \quad (14)$$

The indirect flux into the pole can be represented as the sum of the “ideal” uniform indirect flux into each rectangular pole surface, plus the excess flux associated with all the pole edges, giving

$$\Psi_i = V_0 \sum \{ \mu \text{Area/gap} \} \quad (15)$$

$$+ V_0 \sum \{ \text{excess flux coeff.} \times \text{pole edge length} \}$$

For a given geometry,  $\Psi_d + \Psi_i = 0$  gives  $V_0$ , and with it the fields. In many circumstances, the poleface dimensions are set by specifications determined by the application of the magnet, and one has to design the height of the pole to get the desired field level. It is clear that both  $\Psi_d$  and  $\Psi_i$  are linear functions of that pole height, making this analytical design procedure simple.

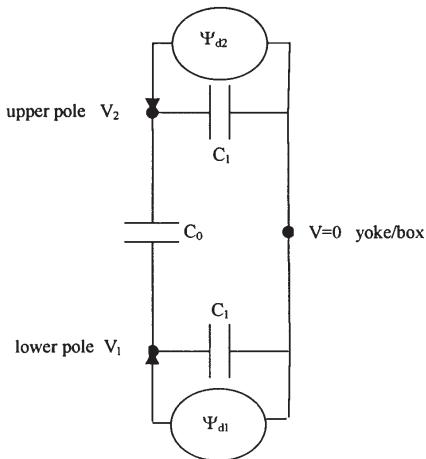


Figure 6: Capacitor network for box dipole.

Representation of a box dipole by capacities: Since the upper and lower pole have indirect flux flowing to each other, and each pole has flux flowing to the “box”, the equivalent capacitor network shown in Fig.6 depicts these conditions. For normal symmetric direct flux deposition ( $\Psi_{d1} = -\Psi_{d2}$ ), the capacity appearing to the right of  $V_0$  in Eq.(15) equals  $2C_0 + C_1$ . For asymmetric ( $\Psi_{d1} = \Psi_{d2}$ ) excitation (that leads to skew multipoles!), only  $C_1$  appears in the equations governing  $V_1$  and  $V_2$ .  $C_1$  is most easily obtained by calculating the indirect flux for that excitation pattern.

#### Other hybrid devices

*Hybrid insertion devices:* Fig.7 shows a cross-section of a hybrid insertion device for the generation of synchrotron radiation or related applications. Following the procedure described above, the basic design is straight forward, but the execution is more complicated because in these devices, some of the iron is driven into saturation. Indirect flux from the poles in the 3rd dimension can significantly degrade performance. CSEM blocks can be added at these locations, thereby making the side-effects enhance, rather than degrade, the on-axis field strength.

*Solenoidal field focusing magnet:* Fig.8 shows schematically the structure of a hybrid magnet that produces solenoidal fields.

*CSEM assisted EM:* Fig.9 shows the cross-section of one of a quadrupole whose range of operation was extended by the introduction of

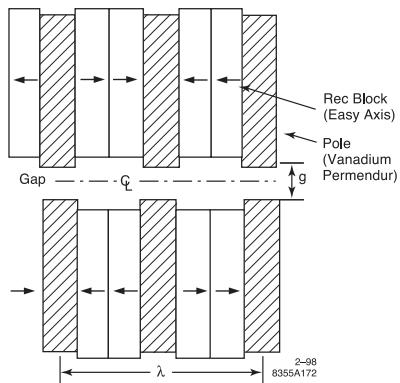


Figure 7: Hybrid insertion device.

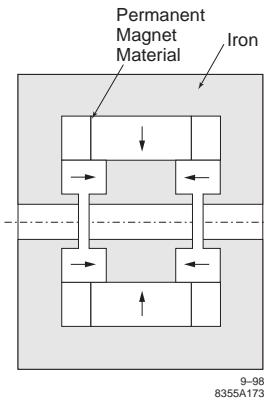


Figure 8: Hybrid solenoid focusing magnet.

CSEM. The CSEM does not contribute directly to the excitation of the magnet, but increases the upper limit of the field dramatically by reverse-biasing the iron pole with PM flux so as to delay EM flux saturation to that corresponding to a higher level of EM excitation. Looking at the magnet without excitation by the coils, it is clear that the iron is “pre”-loaded with a lot of flux generated by the CSEM. If the polarities are chosen right, this will reduce the flux density in the iron at the peak excitation of the magnet, increasing the iron saturation-limited field level that can be reached. The benefits obtained by this use of CSEM are related only to the iron saturation; the Ampere turns required to reach higher field have to be delivered by the power supply. The type of magnet shown in Fig.10 is a typical magnet

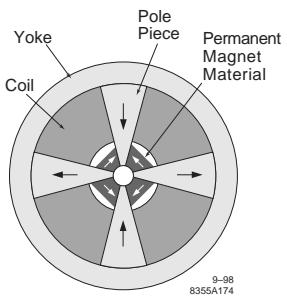


Figure 9: CSEM assisted EM quadrupole.

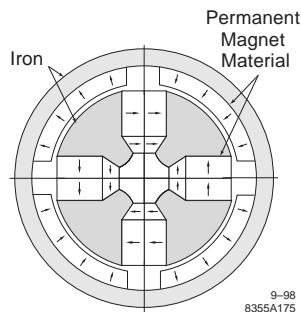


Figure 11: Variable quadrupole.

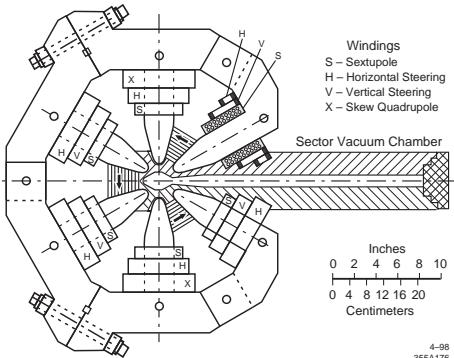


Figure 10: Storage ring sextupole magnet.

which can benefit from this technology. Because of other constraints and performance specifications on synchrotron radiation storage ring magnets, the coils have to be at an unusually large distance from the aperture, leading to severe saturation effects that can be ameliorated by introduction of the blocks of CSEM placed between some of the poles.

### Field quality and magnet fabrication

Hybrid magnets Since in a properly designed hybrid magnet the field distribution is controlled by the iron configuration, the methods developed to design good EMs are directly applicable. There are, however, aspects to the design of hybrid magnets that do not exist in EMs:

(i) When designing EMs it is sometimes acceptable for the overall system to make the pole face wider than absolutely necessary. When doing this in hybrid magnets, one pays an additional price beyond the increase in the weight and cost of the (often not very expensive) iron: the amount of (usually costly) CSEM also increases.

(ii) Due to their intrinsically symmetrical structure and excitation, EMs rarely have skew multipoles. In a magnet as shown in Fig.3, if one of the poles is more strongly excited by the CSEM than the other, skew fields will appear. While these fields decay exponentially with the distance from the edge of a pole, they decay only half as fast as error fields that are perpendicular to the midplane in the midplane. By rotation of the round blocks (four are shown, to increase the tuning range) of CSEM shown in Fig.3, the direct flux deposited onto the two poles can be changed, allowing the scalar potential of both poles to be adjusted to cancel the skew fields and fine-tune the field level precisely. Large tuning ranges can be achieved by magnet configurations as shown in Fig.11, depicting the cross-section of a box-magnet type quadrupole. By rotation of the outer iron ring with the attached CSEM, the field strength can be adjusted over a wide range.

(iii) While the quality of the CSEM used to excite the poles does not have to be very good, placement of CSEM close to the beam requires careful assessment of the effect of CSEM quality on direct fields, and material selection and sorting described below should be used.

(iv) Close to sharp corners of poles similar to those shown in Fig.3, fields can get very large, potentially leading to demagnetization (i.e. driven into the nonlinear part of the magnetization curve) of the CSEM, but fortunately only in a very small volume. This can be avoided by rounding such corners, but usually is not enough of a problem to warrant that extra step.

(v) Demagnetization of the CSEM has to be avoided during all phases of fabrication (starting with careful unpacking of the material) and use of the magnet. This requires good planning because

of the large forces between all components of the magnet.

(vi) Field quality of hybrids is, with equally careful design and fabrication, very similar to the field quality of EMs.

(vii) The local field quality routinely achieved for a typical hybrid insertion device (with a period of 6 cm, a peak field of 1 T, and a length of 300 cm) is  $dB/B = .2\%$ , and the first integral of the field over the length of the magnet is typically  $\pm 100$  G-cm.

Iron free magnets Because of the absence of “equalizing” iron, the demands on the quality of the CSEM are much higher for iron-free magnets than for hybrids. One way to achieve good performance despite imperfect material is the measurement of magnetization uniformity, magnetization strength, and easy axis orientation of the CSEM blocks, and then assigning appropriately the location of each block in the assembly. The condition that the material is never driven beyond the linear part of the magnetization curve is much more difficult to satisfy for iron-free magnets than for hybrids. This becomes clear for a multipole magnet when one considers that for a very large number of blocks in the magnet, the fields outside the magnet are vanishingly small. This means that at some locations on the outer circumference (where one does not expect difficulties at first sight) of the magnet,  $B$  parallel to the easy axis is zero. One has to pay particular attention to demagnetization when one places a multipole inside another magnet, like a detector solenoid. Such a detector field can be very large without causing damage provided the CSEM stays in the linear part of the magnetization curve. This requires very careful alignment and makes it “illegal” to place such a multipole into a fringe field region of a detector with large radial field components.

The field quality of multipole magnets is usually expressed by the relative error of the integral of the field over the length of the magnet; with careful design, sorting, tuning by various means [3], the relative error of this integral can be reduced to the order  $10^{-4}$ . The numbers for iron free insertion devices are similar to those of hybrid insertion devices.

## References

- [1] K. Halbach, NIM 187 (1981) 109
- [2] K. Halbach, NIM 169 (1980) 1
- [3] R.D. Schlueter, Mag. Tech. LBNL Pub 754 LBNL (1995) 55
- [4] R.D. Schlueter, S. Marks, IEEE Trans. Magnetics 32(4) (1996) 2710
- [5] S. Sasaki, NIM A331 (1993) 753

## 7.2.9 Electrostatic Separators

*J.J. Welch, SLAC*

(See also Sec.6.14.)

Electrostatic separators are beamline components that generate strong electric fields perpendicular to orbits in storage rings or synchrotrons. Unlike a magnetic field, a dc electric field applied transversely results in closed orbit distortions which have opposite sign for oppositely charged stored beams. The electric field can therefore be used to separate closed orbits of oppositely charge beams.

**Angular kick** The angular deflection  $\Delta\theta$  of a fully relativistic beam of energy  $E_0$  by a separator with electrodes of length  $L$  and gap  $g$  across which is a total voltage change of  $V$ , is typically

$$\Delta\theta \approx (0.8 - 0.95) \frac{VL}{gE_0} \quad (1)$$

where the numerical factor gets closer to 1 as the ratio of electrode width to gap increases. The angular kick then causes a closed orbit distortion.

**Impedance** Separators present a relatively large impedance to the beam, comparable with a single rf cavity cell. Typical loss factors for separators range from  $k = 0.12$  [4] to  $k = 0.5$  [6, 5] V/pC for approximately 2 cm bunch lengths. This loss generates local rf heating which ranges from a few watts, to a few kW for high current storage rings. Impedance can be reduced by adding ground electrodes in zero potential planes, as well as by tapering and contouring the vacuum chamber outer wall [4]. See Fig.1.

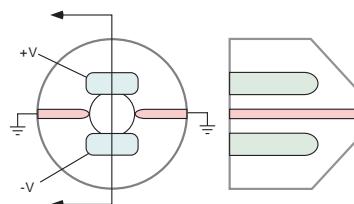


Figure 1: Grounded electrodes emerging from the end of a tapered vacuum chamber help to keep beam electromagnetic fields from being trapped, thereby greatly reducing the loss factor.

**Maximum electric field and voltage** Separators are most often designed to be bipolar with equal and opposite voltage on the two plates. Actual maximum obtainable electric fields and voltage vary considerably in practice and are not accurately predictable. For design purposes the experience of existing separators is a good starting point. See Tab.1. Exotic electrode material and relatively high pressure are used at BNL to get fields of 75 kV/cm at 300 kV.

Table 1: The “maximum operating field” is the maximum operating gap voltage divided by the gap width. Processing voltages and fields are higher. Gaps at CERN are variable from 60 to 160 mm.

| Site     | Operating  | Operating Gap |
|----------|------------|---------------|
|          | Field Max. | Voltage Max.  |
|          | [kV/cm]    | [kV]          |
| CERN [2] | 25         | 400           |
| KEK      | 30         | 240           |
| CESR     | 25         | 250           |
| FNAL     | 50         | 250           |

Table 2: Field limits for various voltages based Kilpatrick criterion for typical separator voltages [6].

| Voltage [kV]  | 120 | 150 | 240 | 300 |
|---------------|-----|-----|-----|-----|
| Field [kV/cm] | 95  | 89  | 80  | 75  |

Design gap fields are well below the Kilpatrick criterion. See Tab.2 (Sec.6.13). Careful electrode cross-section design should limit field enhancements to around 1.5 times the average gap field. Field enhancements at the ends of the electrodes are most significant and may exceed Kilpatrick criteria.

**Synchrotron radiation** SR can generate copious amounts of photocurrent from electrodes causing a high load on the power supplies and local heating, as well as other problems. Photocurrent can be reduced in vertical separators by slightly unbalancing the bipolar voltages so that the zero potential plane is not coincident with the plane where the radiation strikes. Horizontal separators are often built with gaps in the electrodes in the horizontal plane. SR striking the insulating surface of a vacuum feedthrough for negative voltage can also cause beam induced breakdown [3]. This can be avoided by operating in a unipolar (+HV) mode with twice the feedthrough voltage. Scattered SR is important.

**Electrode cooling** Electrodes are often, but not always, actively cooled to prevent warpage as

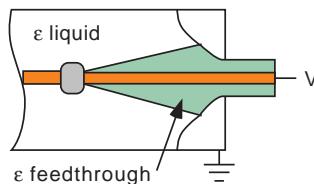


Figure 2: High dielectric constant insulating fluids can cause reduction in the maximum voltage a feedthrough can withstand because it enhances the electric field on the inner conductor where it is the highest.

well as excessive temperature. In order to minimize the risk to the vacuum system of the accelerator, coolants should be vacuum compatible, meaning that they evaporate cleanly and completely in 1 day or less. Cost is also important if substantial quantities are needed. Dielectric strength must be high, conductivity very low, and dielectric constant should not be too high. Conductivity causes excess drain on high voltage power supplies as well as electrolytic corrosion. Conductivities  $\gtrsim 1 \text{ G}\Omega\text{-cm}$  are needed.

High purity water is not really very practical compared with dielectric fluids. High resistivity water,  $> 10 \text{ M}\Omega\text{-cm}$ , is corrosive, so only very long thin circuits can be used to cool the electrodes. Water cannot be used to insulate feedthroughs because of the large surface area, which means a dielectric fluid will be needed in any case. The resistivity of water must be actively controlled with relatively expensive equipment requiring regular maintenance.

At the boundary between the insulating fluid and the feedthrough, since  $\epsilon E$  is continuous, a higher dielectric constant in the fluid lowers the electric field in the fluid and raises the electric field on the inner conductor where it normally is the highest, and can therefore cause feedthrough failure at lower voltage. See Fig.2.

**Radiation safety** With gap voltages well over 100 kV on a vacuum insulated gap, separators will generate large quantities of penetrating x-rays through bremsstrahlung of field emission electrons striking the anode or ground, and are an acute radiation safety problem when energized. Shielding effectiveness of Pb for radiation emitted by an electron beam striking an anode [7] is shown in Tab.4.

Typical and potential fluids are listed below. Ozone depleters, e.g. CFC-113, have been phased out of production but are still in use. Compounds

## Sec.7.2: CONFINEMENT AND FOCUSING

Table 3: Selected properties of dielectric liquids potentially suitable for use with separators. Two different values of dielectric strength of HFE 7100 have been reported.

| Name               | Units     | CFC113 | PF5050 | PF5060 | PF5070 | FC72 | FC77   | HFE7100 | HFE7500 |
|--------------------|-----------|--------|--------|--------|--------|------|--------|---------|---------|
| Manufacturer       |           | Dupont | 3M     | 3M     | 3M     | 3M   | 3M     | 3M      | 3M      |
| Boiling Point      | [°C]      | 48     | 30     | 56     | 80     | 56   | 97     | 60      | 130     |
| Vaporiz. Heat      | [cal/g]   | 35     | 21     | 21     | 21     | 21   | 20     | 30      | 21      |
| Specific Heat      | [cal/g°C] | 0.22   | 0.26   | 0.25   | 0.26   | 0.25 | 0.25   | 0.28    | 0.27    |
| Kinematic          | [cp]      | 0.68   |        |        |        | 0.4  | 0.8    | 0.61    | 0.77    |
| Viscosity 25°C     |           |        |        |        |        |      |        |         |         |
| Dielectric Streng. | [kV/in.]  | 35     |        | 35     | 35     | 40   | 20(35) | 40      |         |
| Dielectric Const.  |           | 2.41   |        | 2.0    | 2.0    | 1.76 | 1.86   |         | 5.8     |
| ODP(CFC11 ≡ 1)     |           | 0.8    | 0      | 0      |        |      |        | 0       | 0       |
| GWP(100 y ITH)     |           | 5000   |        |        |        |      | 480    | 210     |         |
| Rough Cost         | [\$/kg]   | NFS    | 30     | 30     | 30     | 60   | 60     | 36      | 62      |

Table 4: Half Value and Tenth Value thickness of Pb shielding required for radiation emitted by bremsstrahlung from electron beams of different energies.

| Beam Voltage | HVL  | TVL  |
|--------------|------|------|
| [kV]         | [mm] | [mm] |
| 100          | 0.24 | 0.8  |
| 150          | 0.29 | 0.95 |
| 200          | 0.48 | 1.6  |
| 250          | 0.9  | 3.0  |
| 300          | 1.4  | 4.6  |
| 400          | 2.2  | 7.3  |
| 500          | 3.6  | 11.9 |

with high Global Warming Potential (GWP) are on the EPA's last resort list and can only be used in the US if nothing else can do the job. HFE 7100 and HFEE 7500 are CFC replacement products. Note the relatively high dielectric constant of HFE 7500.

**Electrodes** Electrodes are most often made from good grades of SS, although Ti alloy and OFE copper have also been used. The best form of SS seems to be Class 2B sheet with no special surface treatments other than to smooth over welds or burrs with nylon abrasives.

**Insulating vacuum feedthroughs** Perhaps the single most important design feature for vacuum insulators is to adequately shield the “triple junction” where metal simultaneously meets vacuum and ceramic. Without good shielding, electrons can begin “hopping” and multiplying along the insulator surface, resulting in a buildup of positive charge on the insulator which might lead to flashover. The choice of insulating material is usually some form of fairly pure  $\text{Al}_2\text{O}_3$  [1].

Table 5: Typical spark rates per separator per day of operation at various laboratories. H,V refer to horizontal,vertical respectively.

| Site     | polarity | spark rate             |
|----------|----------|------------------------|
|          |          | [sparks/day/separator] |
| CESR (H) | bipolar  | 0.05                   |
| FNAL     | bipolar  | 0.002                  |
| CERN (H) | positive | 0.02                   |
| CERN (H) | negative | 100                    |
| CERN (V) | bipolar  | 0.01                   |
| KEK      | bipolar  | < .03                  |

**Spark rate** Sparks during operation often, if not usually, cause beam loss. Spark rates vary greatly among separators at different accelerators and are generally higher where there is more SR (Tab.5). One cause of beam induced spark has been identified and results from SR striking the insulator of the negative electrode [3].

## References

- [1] R.V. Latham, High Voltage Vacuum Insulation: The Physical Basis, Academic (1981) p.229
- [2] W. Kalbreier et al, EPAC88 or CERN SPS/88-20 (ABT)
- [3] N. Garrel et al, Proc. 2nd Int. Conf. on Space Charge in Solid Dielectrics (CSC'2) (1995), or CERN SL/95-18 (BT)
- [4] J.J. Welch, Z.X. Xu, PAC91, p.1851
- [5] J.N. Weaver, P.B. Wilson, J.B. Styles, SLAC-PUB 2284, PEP Note 287 (1979)
- [6] T. Shintake et al, KEK Report 88-17 (1989)
- [7] NBS Handbook 114, ANSI N543-1974, General Safety Standards for Installations Using Non-Medical X-Rays and Sealed Gamma-Ray Sources, Energies up to 10 MeV

## 7.2.10 Deflection and Crab Cavities

*K. Akai, KEK*

*M. Tigner, Cornell*

### 7.2.10.1 Introduction

In addition to the use of microwave cavity structures, standing or traveling wave, for acceleration, very similar structures can be used for deflecting charged particle beams. Three applications have been in practice for some time: i) time dependent deflection for beam separation in particle physics experiments [1]; ii) time dependent deflection for time to position conversion for short bunch distribution function measurements [2]; iii) so called crab cavities for converting collisions of beams with principal orbits crossing at an angle to head on collisions in the bunch frame. More recently crab cavities are being developed for exchanging, in a storage ring light source, longitudinal bunch extension with vertical extension for producing short x-ray bursts [3].

### 7.2.10.2 Multicell deflection cavities

The design and optimization of deflectors [1] bears many similarities with that for accelerating cavities for electron linacs but there are a few important differences. Deflecting modes for  $v_p = c$  are hybrid TM, TE modes [4] which can be of the forward or backward type. It is important to choose the dispersion curve so that only one type exists in the chosen operating region. The deflection mode polarization must be controlled by use of rods or elliptical cavity shaping. The frequency difference between 0 and  $\pi/2$  orientation must be enough that the operating frequency will excite only one of them. The deflection force although exerted by both electric and magnetic fields is usually given and an equivalent  $E_{\perp}$  [MV/m] with a shunt impedance  $R \equiv E_{\perp}^2 / (-dP/dz)$  [ $\Omega/\text{m}$ ],  $P$  being the total power flow. For  $v_p = c$  the deflection force  $F_{\perp}$  is uniform and aberration free over the iris aperture. In a deflector with attenuation  $\alpha$  and no phase slip the maximum transverse energy kick is

$$p_{\perp}c = eE_{\perp}(0)[(1 - \exp -\alpha \cdot l)/\alpha]$$

for a deflection angle  $p_{\perp}/p$ .

Normal conducting structures have been built for microsecond pulses. A typical example of the genre [2] operates at 2.9 GHz, has a length of 3.6 m, and aperture of 45 mm and a filling time of

0.6  $\mu\text{s}$  and when driven at 20 MW achieved transverse energy kicks of 26 MV. Fields are limited by electric breakdown (see Sec.6.13). Multicell SC structures can operate CW and achieve deflection fields with todays technology of 7 MV/m [5].

### References

- [1] H. Lengeler in Handbook of Accelerator Physics and Engineering, 3rd printing, 1st ed World Scientific (2006) 521
- [2] M. Röhrs, C. Gerth, H. Schlarb, FEL 2006, BESSY, Berlin, 300
- [3] A. Zholents et al, NIM A 425 (1999) 385
- [4] B. Montague in P. Lapostolle and A. Septier, Linear Accelerators, North Holland (1970) 572
- [5] N. Solyak et al, LINAC2004 Lübeck (2004) 797

### 7.2.10.3 Crab cavity

**Crab crossing** When the bunch center passes through an RF deflecting cavity at the zero-crossing phase, the bunch head and tail are deflected in opposite directions by the time-dependent transverse kick, so that the bunch is tilted with respect to the beam trajectory, while the bunch center experiences no deflection. By applying this method in a finite-angle crossing collider, the beams can collide head-on in the center-of-mass frame, while the bunch centroids still cross at an angle. This crossing scheme is called crab crossing, and the deflecting cavities for this use are called crab cavities.

Crab crossing was first proposed for linear colliders [1]. It has been also shown that synchro-betatron coupling terms associated with the crossing angle can be cancelled by applying the crab crossing for storage ring colliders [2]. The finite-angle crossing scheme has the following advantages: no need for separation dipole magnets and much simpler interaction region hardware than that of a zero-angle collision scheme, quick separation after a collision so that parasitic collisions or synchrotron radiation at the detector components are of much less concern. Crab crossing eliminates harmful beam-beam effects caused by finite-angle crossing, and can improve luminosity, while keeping the advantages of the finite-angle crossing scheme.

There are two possible crab crossing schemes in storage rings: one is the so-called local crab crossing (Fig.1, left), where the crab cavities are located before and after the interaction point (IP)

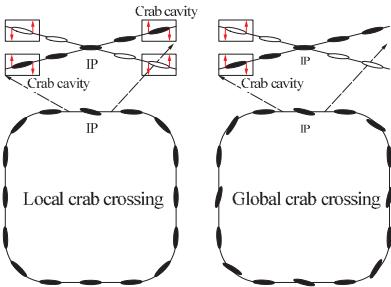


Figure 1: Local crab crossing (left) and global crab crossing (right).

at the betatron phase advance of  $(1/2+\text{integer})\pi$  from the IP. In this scheme bunches are tilted in a crab cavity before the IP, and collide head-on at the IP. The tilt is kicked back to the original orientation by another crab cavity located on the other side of the IP. Thus bunches are tilted only in the region between the crab cavities with no tilt outside this region. Another scheme is termed global crab crossing (Fig.1, right), where a minimum of one crab cavity is located without other crab cavities to remove the tilt. In this scheme the bunch head and tail have different closed-orbit distortions in opposite directions, and thus the bunch is tilted in the whole ring.

The required transverse deflecting voltage for the local crab scheme is given by

$$V_{\perp} = \frac{cE \tan \varphi}{\omega_{rf} \sqrt{\beta_x^* \beta_{x,crab}}}, \quad (1)$$

where  $\beta_x^*$ ,  $\beta_{x,crab}$ ,  $E$ ,  $\varphi$ ,  $\omega_{rf}$  and  $c$  are the beta-function at the IP, the beta-function at the crab cavity, the beam energy, the half crossing angle at the IP, the RF frequency of the crab cavity and the speed of light, respectively. Here, we assume that the beams collide with a horizontal crossing angle. In the global crab scheme, the required transverse deflecting voltage is given by

$$V_{\perp} = \frac{cE \tan \varphi}{\omega_{rf} \sqrt{\beta_x^* \beta_{x,crab}}} \times \frac{2 \sin \pi \nu_x}{\cos(\pi \nu_x - |\Delta \Psi_x^c|)} \quad (2)$$

where  $\nu_x$  is the betatron tune and  $\Delta \Psi_x^c$  is the phase difference between the crab cavity and the IP.

**Crab cavity applications** The first crab crossing was conducted at KEKB in 2007, and crab crossing operation for physics continued until the shutdown in 2010 for machine upgrade. Two superconducting crab cavities, one for the low energy ring (LER) and the other for the high energy ring (HER), were used to make the global

crab crossing. The beam-beam tune shift was improved to 0.09 from the previous 0.056 without crab cavities, although it was lower than the value expected by beam-beam simulations. The crab cavities have been operated successfully at beam currents of 1.8 A in the LER and 1.25 A in the HER.

Interest in crab cavities has been increasing for a wide number of accelerators. Crab crossing is considered as one of possible paths for the LHC upgrade, for its potential for luminosity gain and leveling. Crab cavities for ILC and CLIC are also being investigated to recover luminosity loss due to the crossing angle at the IP. Another application of crab cavities is for generation of short pulse X-ray in storage-ring based light sources. When the tilted bunch passes through an undulator, it radiates X-rays which are also tilted. By slicing out the head and tail part of the X-rays, short pulse X-rays can be obtained [3]. This application is being investigated at the Advanced Photon Source (APS) and the SPring-8 storage ring.

**Crab cavity design** Crab cavities are operated in a deflecting mode that has high transverse shunt impedance. Cavity type can be either standing wave or traveling wave, normal-conducting or superconducting. Design choices and optimization of a crab cavity depends on various factors: required kick voltage, stored beam current and required damping property for parasitic modes, allowable space in both longitudinal and transverse directions, and so on. Several types of crab cavity design for a variety of accelerators have been proposed and studied, including single cell or multi cell cavities operating in the TM110 mode and half wave standing wave TEM resonators.

Crab cavities for high-current accelerators, such as a factory collider, must be a damped structure where parasitic mode impedances are sufficiently reduced to avoid coupled-bunch instabilities. In the case of accelerating cavities, the operating mode (accelerating mode) is usually the lowest frequency mode, and any parasitic mode is a higher-order mode (HOM). Waveguides attached on the cell or enlarged beam pipes whose cut-off frequency is higher than the fundamental mode, but lower than the lowest HOM frequency, can be adopted for heavily damped structures. In the case of crab cavities, on the other hand, lower-order modes (LOM) and same-order modes (SOM) exist as parasitic modes, in addition to the HOM's. The LOM corresponding to

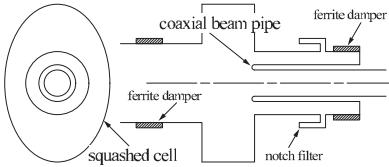


Figure 2: A schematic view of damping scheme for the KEKB crab cavity.

the accelerating mode usually has a high longitudinal coupling impedance, and can give rise to a strong coupled-bunch instability. The unwanted polarization of the crabbing mode as well as some of TE-like mode frequencies can be close to the crabbing mode. Measures are needed to sufficiently damp the parasitic modes including the LOM and SOM, which tends to make the crab cavity design more complicated.

As an example, a schematic view of the damping scheme of the KEKB crab cavity [4] is shown in Fig.2. It has a squashed cell with a coaxial beam pipe attached to the cell. The LOM is coupled monopole-like to the coaxial beam pipe, and propagate in the coax pipe in the TEM mode. The squashed cell makes a large separation of frequencies between the wanted and unwanted TM110 modes, so that the unwanted mode has a higher frequency than the TE11 cut-off of the coax, which makes it propagate in the coax, while the wanted mode has a lower frequency than the TE11 cut-off, and is confined in the cell. The parasitic mode power is absorbed by the ferrite dampers at the end of the coax. In case a small portion of the wanted mode power couples monopole-like due to misalignment of the coax and thus propagate in the TEM mode, it is reflected back to the cell by a notch filter. The other side of the beam pipe is enlarged to extract HOM's.

For the LHC upgrade case, due to limited space between the two beam axes and long bunches of 7.5 cm, several types of compact cavity design at the maximum RF frequency of about 800 MHz are being studied [5]. In the case of ILC, multi cell structures are favored.

**Beam-loading in crab cavities [6]** Since the crabbing mode is a transverse mode, the beam loading on a crab cavity depends on the beam orbit. The loaded Q value of crab cavity should be determined taking tolerance on the orbit, allowed phase shift and available power into

account. Here the transverse shunt impedance ( $\bar{R}_\perp$ ) is defined as

$$\bar{R}_\perp \equiv \frac{V_{\perp c}^2}{P_c}, \quad (3)$$

where  $V_{\perp c}$  is the kick voltage and  $P_c$  is the power dissipated on the cavity surface. (Another notation,  $R_\perp$ , which is usually used in instability calculations, is related as  $R_\perp \equiv \frac{k}{2} \bar{R}_\perp$ , where  $k$  is the wave number.)  $k = \frac{\omega}{c}$

We now consider the case of horizontal crossing. We assume that bunches pass parallel to the cavity axis, displaced by  $\Delta x$  from the central axis. In the case where the bunch spacing is much smaller than the filling time of the cavity, the superposition of the successive single-bunch beam loadings gives a continuous beam-induced voltage ( $V_{\perp b}$ ),

$$V_{\perp b} = V_{\perp br} \cos \psi e^{j\psi}, \quad (4)$$

where

$$V_{\perp br} = -j \frac{I_b \bar{R}_\perp}{1 + \beta} k \Delta x, \quad (5)$$

$$\tan \psi = 2Q_L \frac{\omega_a - \omega_{rf}}{\omega_a}. \quad (6)$$

Here,  $I_b$  is the total beam current,  $\beta$  the input coupling,  $\psi$  the tuning angle,  $\omega_a$  the resonant angular frequency and  $Q_L$  the loaded-Q value.

In the following, a positron beam ( $I_b > 0$ ) is considered. (An electron beam can be treated similarly.) Fig.3 shows a vector relation for the crabbing mode. In this figure,  $V_{\perp g}$  is the generator voltage,  $V_{\perp gr}$  the generator voltage on resonance,  $\alpha_L$  the angle between  $V_{\perp gr}$  and  $V_{\perp c}$ , and  $\phi_c$  the angle of  $V_{\perp c}$  with respect to the beam. It is similar to that for the accelerating mode, except for the beam-induced voltage,  $V_{\perp br}$  and  $V_{\perp b}$ . The phase of  $V_{\perp br}$  with respect to the bunch phase is 90 or 270 degrees, according to  $\Delta x < 0$  or  $\Delta x > 0$  (180 degrees for the accelerating mode). Furthermore, the amplitude of  $V_{\perp br}$  is dependent on  $\Delta x$ .

From the vector relation we obtain

$$\tan \alpha_L = \frac{\tan \psi - Y \cos \phi_c}{1 + Y \sin \phi_c}, \quad (7)$$

$$|V_{\perp gr}| \cos \alpha_L = |V_{\perp c}| (1 + Y \sin \phi_c), \quad (8)$$

where  $Y \equiv \pm |V_{\perp br}/V_{\perp c}|$  (positive sign for  $\Delta x > 0$  and negative sign for  $\Delta x < 0$ ). Since  $|V_{\perp gr}|$  is related to the input power ( $P_g$ ) as

$$|V_{\perp gr}| = \frac{2\sqrt{\beta}}{1 + \beta} \sqrt{\bar{R}_\perp P_g}, \quad (9)$$

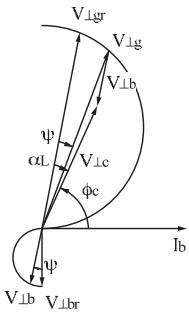


Figure 3: Vector relation for beam loading on the crabbing mode (case  $\Delta x > 0$ ).

the required power to maintain the crabbing voltage is obtained from Eqs. 5, 8 and 9 as

$$P_g = \frac{(1 + \beta)^2}{4\beta\bar{R}_\perp} \times \left\{ \frac{1}{\cos \alpha_L} \left( |V_{\perp c}| + \frac{I_b \bar{R}_\perp}{1 + \beta} k \Delta x \sin \phi_c \right) \right\}^2 \quad (10)$$

## References

- [1] R. B. Palmer, SLAC-PUB-4707 (1988).
- [2] K. Oide and K. Yokoya, PR A40, 315 (1989).
- [3] A. Zholtens et al., NIM-A 425, 385 (1999).
- [4] K. Akai et al Int. J. Modern Phys. A, 2B (Proc. 15th Int. Conf. on High Energy Accel.) (1992) p757
- [5] R. Calaga, R. Tomas, F. Zimmermann, IPAC10, p1240
- [6] K. Akai, Y. Funakoshi, EPAC96, p 2118

### 7.2.11 Electrostatic Lenses

*A. Faltens, LBNL*

**Principles and limitations [1, 2, 3, 4, 5]** Electrostatic fields are used for focusing in injectors, and for transport of ions at low energies. For given charge,  $q$ ,  $E$  and  $B$  fields, the Lorentz force  $\vec{F} = q(\vec{E} + c\vec{\beta} \times \vec{B})$  will be stronger for the  $E$  field at sufficiently low  $\beta$ . The  $\beta$  where the forces are equal depends on the type and size of the lens desired and on the attainable field strengths for it. For reliable operation, the electric fields in vacuum are limited to about  $10^7$  V/m by field emission for distances less than  $10^{-2}$  m. For longer distances other mechanisms decrease the operating fields approximately as  $d^{-1/2}$  up to about the 1 MV voltage level, beyond which data are

scarce. Safe operating voltages should be kept to roughly half of the breakdown voltages, which depend upon materials, surface finish, contaminants, and conditioning. They are often limited by the supporting insulators, which should be shielded. Voltage holding tends to increase with spark conditioning for energies in the spark less than 1 Joule, and deteriorate for energies much higher than that. See also Sec.6.14.

At an electric field of  $10^7$  V/m and an iron-dominated quadrupole magnet with a pole tip field of 1 T the forces are equal at  $\beta = 1/30$ . Scaling of multipole designs to small dimensions is favorable for electric and permanent magnet lenses, and unfavorable for conventional and superconducting magnet lenses. For electric quadrupoles, the electric field gradient  $dE/dx$  can increase as  $1/r$  or faster as the quadrupole is made smaller. The electrodes also provide good electric shielding of beam fields, so multiple beam arrays can be utilized for transport of intense beams. The benefits of an array of small beams are balanced by the cost of constructing small channels precisely.

**Implementation** Electric lenses can be the result of the radial fringing fields of an axisymmetric system or of the transverse fields of a nonaxisymmetric system. They produce greater aberrations than similar magnetic lenses in part because particles change their energy as well as direction within them. The fringing fields can be different in electric and magnetic systems because the source of the electric field is a charge distribution on an electrode but whose potential is usually specified, whereas in a magnetic system the source of the field is a current distribution which is specified directly; a consequence of this can be a larger charge buildup at the ends of an electric element and stronger fringing fields. The surrounding enclosure has a direct effect on the fringing fields. Nevertheless, good electrostatic optical elements can be realized.

In an axisymmetric system such as that formed by two coaxial cylinders spaced a distance apart, or the “acceleration gap”, shown in Fig.1, the focusing action arises from the radial fringing fields.

At one end of a long gap of length  $g$  and potential difference  $\Delta V$  across it, there is a focusing action with the focal length

$$f = 4V/E_z \quad (1)$$

where  $V$  is the kinetic energy of a singly charged, nonrelativistic, ion and  $E_z = \Delta V/g$ . At the other

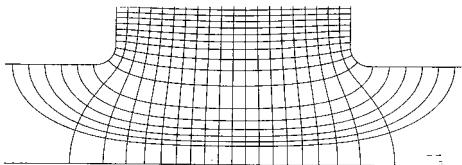


Figure 1: Field and potentials for a “wide gap” cylindrically symmetric lens.

end of the gap there is a defocusing action with

$$f = 4(V + \Delta V)/E_z \quad (2)$$

Both of these strong lenses together therefore yield a weaker net focusing action, whether the gap is an accelerating or decelerating gap. In general, the entrance and exit regions would not be well separated and Eqs.(1,2) would become inadequate. A large body of literature exists for such lenses under the headings of electron optics and electron microscopy [1, 2, 3, 4, 5]. They are not much used in accelerators, with the exception of the gun region. If one imagines a parallel beam generated at the midpoint of Fig.1, and accelerated to the right, then that beam experiences a strong defocusing action as it leaves. The exit lens defocusing and aberrations are primary concerns in gun design. To some extent the analytic design has been superseded by computer codes, such as POISSON and EGUN (Sec.7.1.1.4). The computed electrode shapes are machined to an accuracy of 1 part in  $10^3$  and experience shows that the unwanted multipoles introduced thereby are of the order of 0.1%.

The nonaxisymmetric lenses are analogous to the magnetic multipoles in common accelerator use. The pole shapes which provide good fields for iron dominated magnets are the same shapes for generating electric multipole fields. An important particular example is the hyperbolic pole shape which produces a pure quadrupole field; this is well approximated by a circular cylinder electrode whose radius is 1.1463 times ( $\sim 8/7$ ) the radius of the aperture. Such a quadrupole is shown in Fig.2 for use in focusing an 800 mA  $K^{+1}$  beam at an energy of 2 MeV with electrode potentials of  $\pm 70$  kV.

Transverse multipole fields can also be produced with a large number of small conductors surrounding the beam whose potentials are adjusted to approximate the desired distribution. Such a “squirrel cage” construction is the electrostatic analogue of a current-dominated magnet.



Figure 2: Electrostatic quadrupole formed with cylindrical surfaces.

As before, there are differences at the conductor boundary between the electric and magnetic field cases. There are high spatial-frequency field errors which die off in a distance  $\sim$  the conductor separation. For example using 1 mm W rods, spaced 1 mm apart, it is possible to generate a combined function dipole-quadrupole of good quality [6]. With such close spacings the conductors also shield each other and improve voltage holding.

## References

- [1] J.R. Pierce, Theory and Design of Electron Beams, Van Nostrand-Reinhold (1954)
- [2] P.W. Hawkes, Quadrupole Optics, Springer Tracts in Modern Physics 42, Springer (1966)
- [3] A. Septier ed., The Focusing of Charged Particles, Academic (1967)
- [4] P.W. Hawkes, Quadrupoles in Electron Lens Design, Advances in Electronics and Electron Physics, Supplement 7, Academic (1970)
- [5] A.B. El-Kareh, J.C.J. El-Kareh, Electron Beams, Lenses, and Optics, Academic (1970)
- [6] C.M. Celata et al, Transverse Combining of Four Beams in MBE-4, Fusion Engineering and Design 32-33 (1996) 219

### 7.2.12 Lithium Lens *G. Dugan, Cornell*

**General considerations** The Li lens [1]-[11] is an example of the class, axisymmetric electromagnetic lens. They are used in the collection of high energy secondary particles produced by irradiation of a target by a primary beam. In pulsed operation these lenses can have short focal lengths

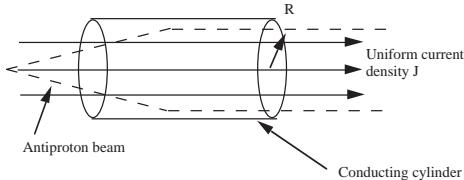


Figure 1: Axisymmetric electromagnetic lens.

Table 1 Candidate lens materials

| Mat'l. | A  | $\sigma_{abs}$<br>barn | $\rho_{lens}$<br>gm/cc | $\ell_{max}$<br>cm |
|--------|----|------------------------|------------------------|--------------------|
| Li     | 7  | 0.16                   | 0.53                   | 13.5               |
| Be     | 9  | 0.2                    | 1.85                   | 4                  |
| Al     | 27 | 0.42                   | 2.7                    | 3.9                |
| Cu     | 63 | 0.78                   | 8.9                    | 1.5                |

and large collecting areas making them candidates for high intensity, low duty secondary beam lines, antiproton and positron sources. Fig.1 illustrates the principal features.

For a uniform axial current density in the lens, the field varies linearly with radius  $r$ ,

$$B = \mu_0 J r / 2 = G r; \quad f [m] = p / 0.3 G \ell \quad (1)$$

where  $f$  is the focal length,  $G$  is the field gradient [T/m],  $p$  is the momentum of the secondary particles [GeV/c] and  $\ell$  the length [m]. If  $\theta$  is the largest production angle to be collected and  $R_0$  the lens radius,

$$\theta = R_0 / f; \quad G = \theta p / 0.3 R_0 \ell [T/m] \quad (2)$$

A large  $\ell$  will minimize  $G$  but  $\ell$  is limited by absorption of the secondaries in the lens,

$$\ell < 0.167 A_{lens} / \sigma_{abs} \rho_{lens} [\text{cm}] \quad (3)$$

for 10% absorption or less.  $A$  is the atomic number,  $\sigma$  the absorption cross section [barn] and  $\rho$  the target density [gm/cc]. Tab.1 displays candidate lens materials.

As an example (the FNAL  $\bar{p}$  source)  $R_0 = 1$  cm and the collection angle is 50 mrad giving  $f = 20$  cm,  $GR_0 \approx 10$  T and, for  $p = 8.9$  GeV/c,  $I = 2\pi R_0^2 G / \mu_0 = 500,000$  [Amp]. Pulsed operation is required.

**Excitation** The lens is excited by placing it in the secondary circuit of a transformer [2, 3] as shown in Fig.2. The transformer steps up the current by the turns ratio  $n$ . The primary resistance is  $R_p$ , the ac resistance and inductance of the lens are  $R$  and  $L$ .

When the SCR switch is closed the capacitor  $C$  charged to potential  $V$ , discharges through the

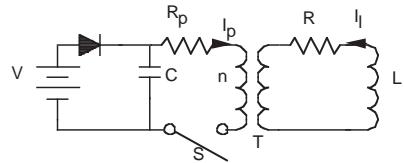


Figure 2: Lens excitation circuit.

 Table 2 Optimum phases for the Li lens,  $\gamma = 0$ 

| $\eta$ | $\chi_{opt} (\text{rad.})$ | $\Delta(\chi_{opt}) \%$ |
|--------|----------------------------|-------------------------|
| 0.3    | 2.94                       | 3.4                     |
| 0.4    | 2.42                       | 2.05                    |
| 0.5    | 2.14                       | 1.09                    |
| 0.6    | 1.98                       | 0.58                    |
| 0.7    | 1.87                       | 0.33                    |
| 0.8    | 1.80                       | 0.20                    |
| 0.9    | 1.75                       | 0.13                    |
| 1.0    | 1.72                       | 0.08                    |

transformer. The resulting secondary current  $I_l$  flows through the lens. After  $C$  is discharged,  $S$  opens. For an ideal transformer

$$I_l = I_0 \exp(-\alpha t) \sin \left( \sqrt{\omega_0^2 - \alpha^2} t \right) \quad (4)$$

$$I_0 = \frac{V}{nL\sqrt{\omega_0^2 - \alpha^2}}; \quad \alpha = \frac{R_e}{2n^2 L}$$

$$R_e = n^2 R + R_p; \quad \omega_0^2 = 1/n^2 LC$$

The resulting field, neglecting end effects is [12]

$$\vec{B}(\rho, \chi) = \hat{\phi} B_0 \operatorname{Re} \left[ \frac{J_1(\rho P)}{J_1(P)} i \exp(K) \right] \quad (5)$$

$$P = \sqrt{2} \sqrt{i + \gamma} / \eta; \quad K = (-i + \gamma) \chi$$

$$\chi = \omega t; \quad \gamma = \frac{\alpha}{\omega}; \quad \delta = \sqrt{\frac{2}{\omega \sigma_c \mu}}$$

$$\eta = \frac{\delta}{R_0}; \quad \rho = \frac{r}{R_0}; \quad B_0 = \mu I_0 / 2\pi R_0$$

$\delta$  is the skin depth and the conductivity  $\sigma_c = 6.94 \times 10^4 / \Omega \cdot \text{cm}$  in Li at 20°C while  $\mu$  of Li is  $4\pi \times 10^{-7}$ .  $J_1$  is the Bessel function. This result neglects the initial transient field which is small. Because of current penetration effects associated with the ac excitation, the field is not in general a linear function of radius. This can be quantified by calculating the rms deviation of the field from a pure linear function  $\Delta(\chi)$ . The phase for which this is the smallest is called the optimum phase. Some values of  $\Delta$  and  $\chi_{opt}$  are shown in Tab.2.

Values of  $\eta > 0.5$  are needed for better than 1% linearity. That is, the skin depth should be

greater than half the lens radius, implying the relationship

$$T_{\text{lens}} > \sigma_c \mu \pi R_0^2 / 8 \quad (6)$$

where  $T_{\text{lens}}$  is the duration of the half-sine-wave pulse. Larger radii require longer pulses.

**Linear optics** The field in the lens may be expanded:

$$\vec{B}(\rho, \chi) = \hat{\phi} B_0 \sum_{j \text{ odd}} c_j(\chi) \rho^j \quad (7)$$

$$c_1 = \text{Re} \left[ F \frac{i}{\sqrt{2} J_1(P)} \exp(K) \right]; \quad F = \frac{\sqrt{i + \gamma}}{\eta}$$

$$c_3 = \text{Re} \left[ F^3 \frac{-i\sqrt{2}}{8 J_1(P)} \exp(K) \right]; \quad K = (-i + \gamma)\chi$$

$$c_5 = \text{Re} \left[ F^5 \frac{i}{48\sqrt{2} J_1(P)} \exp(K) \right]; \quad P = \sqrt{2} F$$

Taking the first order field term in the equation of motion we have for the linear lens matrix

$$M(k, \ell) = \begin{bmatrix} \cos \sqrt{k}\ell & \frac{\sin \sqrt{k}\ell}{\sqrt{k}} \\ -\sqrt{k} \sin \sqrt{k}\ell & \cos \sqrt{k}\ell \end{bmatrix} \quad (8)$$

where  $k = 0.3 B_0 c_1(\chi) / p R_0$  [m<sup>-2</sup>] with  $p$  in GeV/c, other units SI. The focal length of the linear lens is

$$f = 1/\sqrt{k} \tan \sqrt{k}\ell \quad (9)$$

For a typical focusing set up see Sec.7.1.5, Fig.4.

**Aberrations** The nonlinear terms in the equations of motion couple the horizontal and vertical motion. If the nonlinearities are relatively small, perturbation calculations give accurate results. End effects need to be included [12] as well.

Multiple coulomb scattering in the lens will cause some emittance growth. This is generally not a large effect because of the low  $Z$ , but needs to be included in the design (see Sec.7.1.5). The scattering from the lens windows can dominate the Li itself. The relative importance of multiple scattering can be assessed by comparing its rms value to that of the characteristic angle of the beam downstream of the target ( $L\theta\sqrt{k} \sin \sqrt{k}\ell$ )/2,  $L$  being the target length (see Sec.7.1.5).

Since the lens is often required to pass a large momentum spread the chromatic bandwidth is important. Variation in momentum will cause a variation in focal length  $\Delta f/f = \Delta p/p$ . Thus, the momentum spread introduces an increase in the angular divergence of the beam at the second principal plane (Sec.7.1.5). If we define the

momentum bandwidth as that value of  $\Delta p/p$  for which the additional divergence is just equal to one half of the intrinsic angular divergence of the beam at this point, then

$$(\Delta p/p)_{\text{max}} = L/4f \quad (10)$$

This favors lenses with short focal lengths. At the FNAL  $\bar{p}$  source,  $L = 5$  cm,  $f = 20$  cm and the bandwidth is  $\pm 12\%$ .

**Electro-mechanical design issues** Major design issues [2, 13] revolve around the support structure (container) for the Li. The structure must: (i) have low electrical conductivity w.r.t. Li; (ii) contain the Li during the high temperature casting of the Li (i.e. it must not react chemically with liquid Li which is very reactive; (iii) contain the Li during operation with its thermal shocks; (iv) provide cooling of the Li between pulses so that the Li never reaches the melting point.

This discussion is limited to solid Li lenses. Liquid lenses have been built but the engineering issues are quite different [6].

In the following, the current is assumed to be uniform in the lens and the container is assumed to be a hollow cylinder of some material surrounding the Li and having thickness  $w \ll R_0$ . Damping of the lens current pulse is ignored. Joule heating results in a temperature rise,

$$\Delta T_{\text{pulse}} = I_0^2 T_{\text{lens}} / 2\pi^2 \rho_{\text{lens}} C_p \sigma_c R_0^4 \quad (11)$$

where  $\rho_{\text{lens}}$  (20°C) is the density, 0.534 gm/cc and  $C_p$  is the specific heat, 3.51 J/gm-K for Li.

If  $\tau$  is the pulse repetition period and  $t_R = \rho_{\text{lens}} C_p w R_0 / 2\kappa_J$  the thermal relaxation time where  $\kappa_J$  is the jacket thermal conductivity, then the max. post pulse temperature is

$$\hat{T} = \Delta T_p / [1 - \exp(-\tau/t_R)] \quad (12)$$

This temperature must be less than 180°C, the melting point of Li. However the condition imposed by the stresses in the container are more restrictive. The max. hoop stress induced in the jacket by the Li expansion is

$$\hat{s} = \alpha \hat{T} M_B R_0 / w (1 + 2M_B R_0/Ew) \quad (13)$$

where  $M_B$  is the bulk modulus of compressibility of Li (1630 kpsi),  $\alpha$  is the volume expansion coeff. of Li ( $1.68 \times 10^{-4}$ ) and  $E$  the Young's modulus of the jacket. For reliable operation over millions of cycles the stress must be less than the million cycle fatigue limit. For the FNAL  $\bar{p}$  source,  $R_0 = 1$  cm, the jacket is of Ti6Al4V,  $w = 0.1$  cm,  $\kappa_J = 0.075$  W/cm-K and  $E = 15,000$  kpsi.  $I_0 =$

## Sec.7.2: CONFINEMENT AND FOCUSING

0.67 MAmp,  $T_{\text{lens}} = 0.35 \mu\text{s}$ ,  $\tau = 2 \text{ s}$ , giving  $\hat{s}$  of 70 kpsi. The fatigue limit for this alloy is 116 kpsi.

The magnetic forces on the Li produce a compressive body force

$$dF/dV = -\mu_0 I^2 r / 2\pi R_0^4 \quad (14)$$

neglecting current penetration effects. To prevent this pinch from separating the Li from the cooling jacket the Li is preloaded during casting. The preload (75-100 Atm.) is designed to exceed the pinch pressure and preserve thermal contact with the jacket. Exact determination of the preload is made using finite element calculations taking the body force and thermal expansion into account.

The Li must be contained axially by windows through which the secondary beam must pass. These windows should be thin to minimize multiple scattering but strong enough to contain the Li preload. Both Be and Ti have been used [14]. The detailed design of the shape is a compromise between structural requirements and desire for minimal scattering.

Fabrication of the lens is done by casting molten Li into the cylindrical volume of the jacket and end caps. The process occurs at about  $200^\circ\text{C}$  just above the Li melting point. The casting is done in vacuo to avoid exposing the Li to air. A hydraulic system coupled to the molten Li establishes the preload.

## References

- [1] B.F. Bayanov, G.I. Silvestrov, Sov. Phys. Tech. Phys., Vol.23, 1 (1978) p.94
- [2] B.F. Bayanov et al, NIM 190 (1981) 2
- [3] G. Dugan et al, PAC 83, p.3660
- [4] B.F. Bayanov et al, Proc. XII Int. Conf. High Energy Acc. (1983) p.587
- [5] R. Bellone, A. Ijspeert, P. Sievers, Proc. XIII Int. Conf. High Energy Acc. (1985) p.272
- [6] B.F. Bayanov et al, EPAC 88, p.263
- [7] S. O'Day, K. Anderson, PAC 96, p.1888
- [8] R. Schultz, P. Hurh, PAC 03, p.1449
- [9] P. Hurh, J. Morgan, R. Schultz, PAC 03, p.1443
- [10] P. Hurh, A. Leveling, PAC 03, p.1440
- [11] P. Hurh, J. Morgan, S. Tariq, PAC 03, p.641
- [12] T.A. Vsevolozhskaya, M.A. Lyubinova, G.I. Silvestrov, Sov. Phys. Tech. Phys. Vol.20, 12 (1976) p.1556
- [13] S. O'Day, K. Anderson, PAC 95, p.1888
- [14] P. Sievers et al, PAC 85, p.3066

## 7.2.13 Orbit Feedback Control

*Y. Tian, L.H. Yu, BNL*

### 7.2.13.1 Principles

(See also Sec.4.7.3) Fig.1 shows the structure of a general orbit feedback system. Here  $\mathbf{R}(s)$  is the response matrix of the accelerator,  $\Theta(s)$  is the corrector strength and  $\mathbf{d}(s)$  is the BPM measurement. Due to vacuum wall eddy currents and other factors in the accelerator, the relationship between  $\Theta$  and  $\mathbf{d}$  depends on frequency. Laplace transforms are used to describe the system in frequency domain ( $s = j\omega$ ).  $\mathbf{e}(s)$  represents the system noise, which includes the contributions from quads center motion, BPM motion, BPM electronics noise, and angular kicks from stray field errors.  $\mathbf{d}_{\text{golden}}(s)$  is the reference orbit, and  $\mathbf{E}(s)$  is the orbit error.

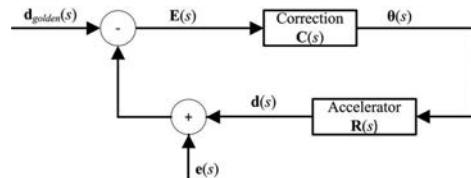


Figure 1: Structure of orbit feedback system.

If the system is a simple single input single output (SISO) system, the orbit error can be written as:

$$\mathbf{E}(s) = \frac{1}{1 + \mathbf{R}(s)\mathbf{C}(s)} [\mathbf{d}_{\text{golden}}(s) - \mathbf{e}(s)] \quad (1)$$

The goal of the orbit feedback system is to measure the  $\mathbf{R}(s)$  and design the corresponding  $\mathbf{C}(s)$  in order to suppress the sensitivity to  $\mathbf{e}(s)$  and get a stable beam orbit. This is a standard SISO problem in control theory [1, 2].

For a real accelerator, the orbit feedback system is a multiple-input and multiple-output (MIMO) control system. For a typical global orbit feedback system, the number of BPMs and correctors can vary from several tens to a few hundreds. For such a MIMO system,  $\mathbf{R}(s)$  and  $\mathbf{C}(s)$  in Eq.(1) are matrices and other terms are vectors.

One can study the stability problem of the MIMO system by looking into the details of the accelerator system. The  $\mathbf{R}(s)$  consists of three parts and can be written as [3]:

$$\mathbf{R}(s) = \mathbf{F}_B(s)\mathbf{R}_0\mathbf{F}_V(s) \quad (2)$$

where  $\mathbf{R}_0$  is the response matrix between corrector magnet fields and beam orbit. When a corrector magnet field changes, the electro-magnetic force applied to beam will change immediately and the beam orbit will also change quickly.  $\mathbf{R}_0$  can be treated as independent of frequency.

In Eq.(2), matrix  $\mathbf{F}_B(s)$  is the BPM transfer function between beam orbit and BPM output. Matrix  $\mathbf{F}_V(s)$  is the transfer function between corrector current setpoints and corrector magnetic fields.

Assuming each BPM has the same frequency response  $f_B(s)$ , the matrix  $\mathbf{F}_B(s)$  is then independent of BPMs and can be written as:

$$\mathbf{F}_B = f_B(s)\mathbf{I} = \frac{\alpha_B}{s + \alpha_B}\mathbf{I} \quad (3)$$

where  $\alpha_B$  is the pole of the transfer function and is typically in the kHz range.  $\mathbf{I}$  represents a unit matrix.

Similarly, if each corrector has the same frequency response  $f_V(s)$ , the matrix  $\mathbf{F}_V(s)$  is then independent of correctors and can be written as:

$$\mathbf{F}_V = f_V(s)\mathbf{I} = \frac{\alpha_V}{s + \alpha_V}e^{-s\tau}\mathbf{I} \quad (4)$$

where  $\alpha_V$  is the pole of the transfer function and is typically in the kHz range.  $\tau$  is the delay in the power supplies, including the effect of the vacuum chamber next to the fast corrector. The communication and computation delay is also included in  $\tau$ .  $\tau$  is typically less than 1ms. For example,  $\tau$  is less than 0.2ms for NSLS-II.

Combining Eq.(2), (3) and (4), one obtains:

$$\mathbf{R}(s) = f(s)\mathbf{R}_0 \quad (5)$$

where

$$f(s) = f_B(s)f_V(s) \quad (6)$$

Eq.(5) shows that the transfer function matrix  $\mathbf{R}(s)$  can be simplified to the multiplication of a single transfer function  $f(s)$  and a constant matrix  $\mathbf{R}_0$ . The singular value decomposition (SVD) [4] algorithm can further help to convert the MIMO problem into an SISO problem. Using the common SVD notation, we assume there are  $M$  BPMs and  $N$  fast correctors in the orbit feedback system. When  $M \geq N$ , the SVD of response matrix  $\mathbf{R}(s)$  can be written as:

$$\mathbf{R}(s) = f(s)\mathbf{R}_0 = f(s)\mathbf{U}\Sigma\mathbf{V}^T \quad (7)$$

where  $\mathbf{U}$  is a  $M \times N$  matrix with orthonormal column vector set.  $\mathbf{V}^T$  is the transpose of the  $N \times N$  matrix  $\mathbf{V}$ , which also has an orthonormal column vector set.  $\Sigma$  is an  $N \times N$  diagonal matrix with

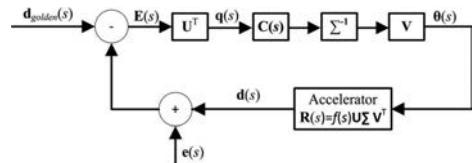


Figure 2: Structure of orbit feedback system with SVD.

non-negative elements and its eigenvectors are referred as the orbit eigenmodes. The  $N$ -dimensional space constructed by these eigenmodes is referred to as eigenspace.

Fig.2 shows the structure of an orbit feedback system with SVD. In Fig.2,

$$\mathbf{q}(s) = \mathbf{U}^T \mathbf{E}(s) \quad (8)$$

where  $\mathbf{q}(s)$  is the orbit error representation in eigenspace. We will show that in eigenspace, each eigenmode is independent and can be treated as an SISO problem. From Fig.2,

$$\begin{aligned} \mathbf{d}(s) &= f(s)\mathbf{U}\Sigma\mathbf{V}^T\Theta(s) \\ &= f(s)\mathbf{U}\Sigma\mathbf{V}^T\mathbf{V}\Sigma^{-1}\mathbf{C}(s)\mathbf{q}(s) \end{aligned} \quad (9)$$

Since  $\mathbf{V}\mathbf{V}^T = \mathbf{V}^T\mathbf{V} = \mathbf{I}$  and  $\Sigma\Sigma^{-1} = \mathbf{I}$

$$\mathbf{d}(s) = f(s)\mathbf{U}\mathbf{C}(s)\mathbf{q}(s) \quad (10)$$

Also from Fig.2,

$$\mathbf{q}(s) = \mathbf{U}^T[\mathbf{d}_{golden}(s) - \mathbf{e}(s) - \mathbf{d}(s)] \quad (11)$$

Therefore,

$$\mathbf{q}(s) = \mathbf{U}^T[\mathbf{d}_{golden}(s) - \mathbf{e}(s) - f(s)\mathbf{U}\mathbf{C}(s)\mathbf{q}(s)] \quad (12)$$

Since  $\mathbf{U}^T\mathbf{U} = \mathbf{I}$

$$\mathbf{q}(s) = \mathbf{U}^T\mathbf{d}_{golden}(s) - \mathbf{U}^T\mathbf{e}(s) - f(s)\mathbf{C}(s)\mathbf{q}(s) \quad (13)$$

In the right side of Eq.(13), the first two terms are the eigenspace representations of golden orbit and noise. The third term is the compensation in eigenspace. In eigenspace, we need to design compensation for each eigenmode, or,

$$\mathbf{C}(s) = \begin{bmatrix} c_1(s) & 0 & 0 & 0 \\ 0 & c_2(s) & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & c_N(s) \end{bmatrix} \quad (14)$$

Then, for the  $i$ th eigenmode,

$$q_i(s) = [\mathbf{U}^T\mathbf{d}_{golden}(s)]_i - [\mathbf{U}^T\mathbf{e}(s)]_i - C_i(s)q_i(s) \quad (15)$$

Or

$$q_i(s) = \frac{1}{1 + f(s)C_i(s)} \times \{[U^T d_{golden}(s)]_i - [U^T e(s)]_i\} \quad (16)$$

Comparing with Eq.(1), one can easily recognize that the goal is to design the SISO compensation  $C_i(s)$  to suppress the noises for the  $i$ th eigenmode. Thus the  $M \times N$  MIMO problem is converted into  $N$  SISO problems in eigenspace.

Even though the above discussions are for continuous system, they are valid for discrete digital systems. Today most of the fast orbit feedback systems are discrete systems. BPM data are digitized and transferred for the feedback calculation. The feedback calculation is carried out by digital computer, DSP or FPGA. The corrector setpoints are also digital signals to power supplies. As it is well-known in control theory [5], the Laplace transform mentioned above should be replaced by Z-transform and the goal is to design  $N$  digital compensations in eigenspace, and Eq.(14) becomes:

$$C(z) = \begin{bmatrix} c_1(z) & 0 & 0 & 0 \\ 0 & c_2(z) & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & c_N(z) \end{bmatrix} \quad (17)$$

Here the  $c_i(z)$  refers to the Z transform of the digital compensator for the  $i$ th eigenmode.

Eq.(15) shows that each eigenmode is independent in eigenspace. This independence can be verified in a real system. Fig.3 shows input signals are injected in one eigenmode and the responses are measured in eigenspace. At first iteration, due to measurement errors, the response of the system may not be accurately diagonal. However, based on the first result one can refine the  $V$  matrix such that the system is more accurately diagonalized [6, 7]. After several iterations, the

response should only appear in the same eigenmode, and responses in other modes are made equal to zero within measurement errors.

Not only can we verify the independence of eigenmodes in Fig.3, but also we can study the system in each eigenmode using standard SISO control theory. For example, the transfer function of the whole accelerator system in each eigenmode can be measured by injecting various input signals into each eigenmode and by measuring the response in eigenspace. The measurements provide both time domain and frequency domain characteristics of the system. Based on the measurements, the compensation  $C_i(s)$  or  $C_i(z)$  for each eigenmode can be designed.

In Fig.3, the noise in eigenspace can be measured if the injected input signals are set to zero. In this way the noise characteristics in both time domain and frequency domain are obtained. The frequency domain response is also called orbit error spectrum [8], which provides useful information about the system noise.

### 7.2.13.2 Implementation

In an orbit feedback system, the typical goal is to suppress noise up to a few hundred Hertz. To achieve this goal, the orbit feedback sampling/correction frequency needs to be in KHz range. For example, NSLS-II's orbit feedback system's sampling/correction frequency is 10 KHz. During one sampling/correction cycle, the system needs to finish BPM data transfer, orbit calculation and power supply setpoints delivery. For a large distributed system, the system architecture needs to be carefully designed to meet this stringent time budget.

In a typical global orbit feedback system, there are many (from a few tens to a few hundred) BPMs, each of them is capable of generating beam position data at 10 KHz. The first challenge is to quickly deliver all BPM data for the global orbit feedback calculation. Due to the large number of BPMs, serial link architecture is more suitable than a tree or star architecture. As one example, Fig.4 shows the NSLS-II orbit feedback architecture. Fig.4(a) shows the architecture of orbit feedback system in one cell. Fig.4(b) shows the 30 cells around the storage ring. In each cell, a BNL-developed serial device interface (SDI) is used to transfer the 6 BPM data to the cell controller. A similar SDI link delivers each cell's BPM data around the ring. It takes less than

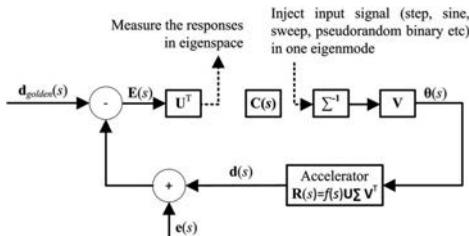
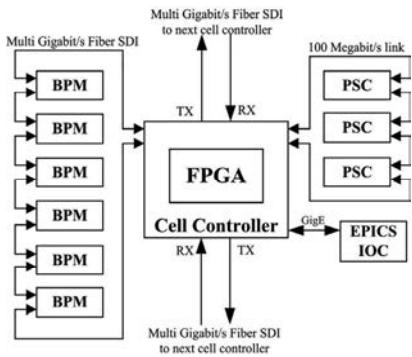


Figure 3: Measure system transfer function in eigenspace.



(a) Structure in one cell.

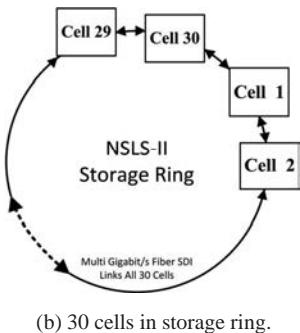


Figure 4: NSLS-II orbit feedback system.

$14\mu s$  to deliver the 180 BPM data (X and Y positions) to each of the cell controllers where the orbit calculation is carried out. The cell controller also has 100Mbit/second serial links to the power supply controllers (PSC) to send the corrector set-points. All the communications are controlled by the FPGA chips in the cell controller, BPM and power supply controller.

For the orbit feedback calculation, the compensation in each eigenmode allows us to study and design the system using standard SISO control theory. On the other hand, this approach requires a large number of calculations [8, 9, 10, 11]. To calculate one corrector setpoint, it takes of the order of  $N \times M$  multiplications and accumulations [9, 10], where  $M$  is the BPM number and  $N$  is the corrector number. For a global feedback system,  $M$  and  $N$  are large and it is difficult to finish the calculation within one feedback cycle ( $100\mu s$  for 10 KHz feedback). This is the second challenge for the system architecture design. A fast computation system is needed to overcome

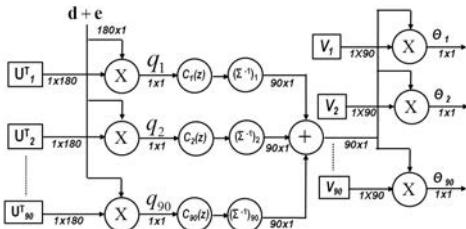


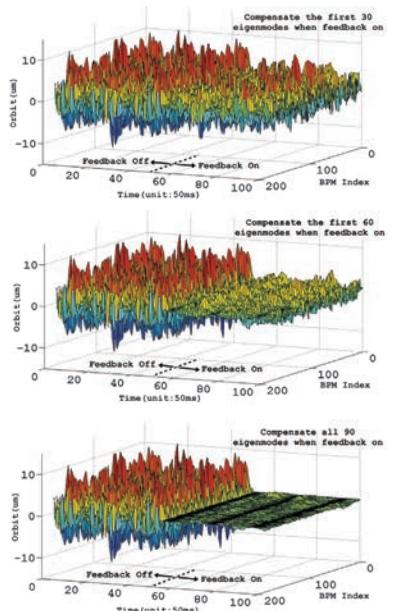
Figure 5: NSLS-II orbit feedback calculation.

this computation challenge. In NSLS-II, a powerful FPGA chip in the cell controller is used for the global orbit feedback calculation. Not only because FPGA has many DSP blocks for high speed calculation, but also because it can carry out the calculations in parallel. The individual eigenmode compensation algorithm can take advantage of the parallel DSP features and thus greatly reduce the amount of calculation time.

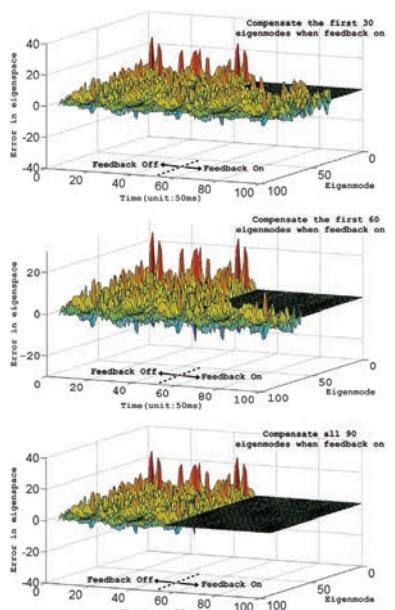
Fig.5 shows the FPGA implementation of NSLS-II global orbit feedback calculation. There are 180 BPMs and 90 fast correctors in the system. It shows the calculation in each cell controller. The eigenspace decomposition and the eigenmode compensation ( $ci(z)$ ) are carried out in parallel for all the eigenmodes. The corrector set-point calculations are also done in parallel. The total calculation time is less than  $10\mu s$ .

Fig.6 shows the simulation results for NSLS-II orbit feedback system. The simulation is based on the NSLS-II lattice and the NSLS-II virtual accelerator. In the simulation, DC to 200Hz random noise is introduced into the BPMs and correctors respectively. Fig.6(a) shows the orbit changes during the feedback off and on period. When feedback is on, different eigenmodes compensations are simulated for comparison. The three figures shows that the orbit gets more stable as the compensated eigenmodes number changes from 30, 60 to 90.

Fig.6(b) shows the orbit error decompositions in eigenspace for the different compensation scenarios. It clearly shows when the first 30 eigenmodes are compensated, the errors in these eigenmodes will be suppressed to nearly zero. Similarly, when the first 60 eigenmodes are compensated, the errors in these eigenmodes become very small. When all the 90 eigenmodes are compensated, not only the errors in eigenspace are all small, but also the orbit becomes more stable, as shown in the last figure of Fig.6(a).



(a) Orbit changes for different compensations.



(b) Errors in eigenspace for different compensations.

Figure 6: Simulation of orbit feedback with different number of eigenmode compensations.

The simulation in Fig.6 confirms that individual eigenmode compensation will suppress the noise in eigenspace and therefore make the orbit more stable. It also shows that the orbit feedback system can be sequentially adjusted in eigenspace. This is one advantage for both machine commissioning and machine study.

### 7.2.13.3 Local orbit feedback

The implementation of a local feedback can be considered similar to that of global feedback with reduced number of BPMs and correctors. For example, each NSLS X-ray insertion device uses two BPMs and four correctors to make a local orbit feedback system [12]. All the above discussions can be applied to the local orbit feedback system.

## References

- [1] R.O. Hettel, IEEE Trans. Nucl. Sci. NS-30(1983) 2228
- [2] L.H. Yu et al, NIM A284 268 (1989)
- [3] L.H. Yu, EPAC08, p.3315
- [4] W. Press et al, "Numerical Recipes" (1992)
- [5] A. Oppenheim et al, "Signals and Systems" (1996)
- [6] O. Singh et al, NIM A418 267 (1998)
- [7] L.H. Yu et al, PAC89, p.1792
- [8] A. Terebilo et al, EPAC06, p.3035
- [9] D. Teytelman, PAC03, p.318
- [10] Y. Tian et al, PAC11, p.1488
- [11] M.G. Abbott et al, EPAC08, p.3257
- [12] O. Singh et al, PAC 91, p.1528

### 7.2.14 Feedback to Control Coupled-Bunch Instabilities

*J.D. Fox, SLAC*

Active electronic feedback systems require i) a beam pickup or sensor, ii) a processing block which calculates a correction signal and iii) correction of beam coordinate via an actuator (kicker) and associated power stages as shown in Fig.1. The correction signal as applied to the beam must be sufficient to give net negative damping  $\gamma_{eff} < 0$  for the desired operating current, and be robust in the presence of external disturbances or other considerations which reduce the effectiveness of the feedback signal.

An unstable system  $G(s)$  can be stabilized via appropriately chosen feedback  $H(s)$ . This feedback topology provides net negative damping, and

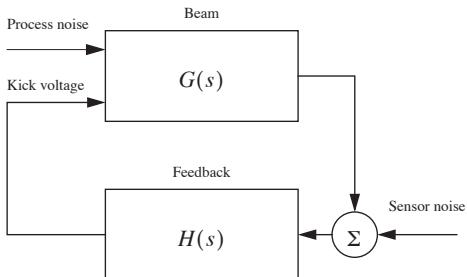


Figure 1: General form of instability control system, with disturbances driving the beam, and a feedback correction applied to the beam. Noise is from within the feedback channel and from external disturbances driving the beam.

acts to reduce process noise by  $1/G(s)H(s)$ . A stable system will still be excited to finite amplitude by external disturbances (e.g. transverse motion from an injection kicker, longitudinal motion from noise in RF accelerating systems). Depending on the system margins and design, driven motion on the beam can saturate the feedback system channel thus reducing the feedback damping term. This can lead to loss of control via exponentially-growing beam motion. A practical system must provide excess damping for a useful margin of stability, but unlimited gain (damping) is not possible for a physical system with time delay which must act over a finite system bandwidth - excess gain causes instability of the feedback loop itself.

**Motion of bunches and time domain/frequency domain representation** The centroid motion of a bunched beam in circular accelerators and storage rings may be represented in the time domain as a harmonic oscillator  $\ddot{x} + 2\gamma_{eff}\dot{x} + \omega_0^2x = U_{ext}(t)$  with oscillation amplitude  $x$ , natural frequency  $\omega_0$ , damping rate  $\gamma_{eff}$  and external driving term  $U_{ext}$ . Fixed impedances interacting with the beam (cavity fundamental and HOM impedances, vacuum chamber resistive wall impedance) provide a driving term  $\gamma_{imp}\dot{x}$  proportional to bunch current \* impedance. For complex external impedances the imaginary part shifts the natural frequency, the real part provides damping or anti-damping depending on the external impedance frequency[1]. Synchrotron radiation losses (if present) provide a damping term  $\gamma_{rad} < 0$ . Above the threshold current for stability  $\gamma_{eff} > 0$  where  $\gamma_{eff} = \gamma_{rad} + \gamma_{imp}$  and

the bunch oscillation amplitude grows exponentially. Below the threshold  $\gamma_{eff} < 0$  motion is damped with envelope of motion  $e^{\gamma_{eff}t}$ . External electronic feedback systems can provide an external driving term proportional to  $\dot{x}$ , and act to increase (or decrease) the system damping.

The behavior of the coupled-bunch system including external impedances can be represented in the time or frequency domain in the normal mode formalism, each mode  $l$  represented by a harmonic oscillator description with modal frequency  $\omega_l$ , effective damping rate  $\gamma_{eff,l}$ . For a uniform fill system with  $M$  bunches, revolution frequency  $F_{rev}$ , there are  $M$  normal modes. The motion of  $M$  modes in the frequency domain will be represented in a frequency spectrum with minimum span of  $M * F_{rev}/2$ . The oscillatory motion will be found in  $\pm\omega_0$  sidebands around  $M/2$  revolution harmonics, with the information folded around RF harmonics repeating in bands in the bands between the RF harmonics spaced every  $M * F_{rev}$  revolution harmonics. For even-spaced and equally-filled bunch systems this decomposition is a Fourier eigenmode decomposition, for arbitrary current distributions or fills with gaps the system normal modes will be a superposition of the even fill eigenmodes.[2]

### Feedback System Implementation Options

A system with a circulating beam, all  $M$  bunches passing sequentially through pickup(s) and kicker(s) presents sequential signals in the time domain (repeating at the revolution frequency  $F_{rev}$ ), and in the frequency domain the pickup information is encoded in a pattern of spectral lines which repeat every  $M * F_{rev}$  frequency interval. Feedback systems are often described as time-domain (bunch by bunch) systems, or frequency-domain (modal) systems.

Frequency-domain processing is practical for a small number of bunches,  $M$ , or a system with a small number of well-defined and known external HOM resonators (e.g small number,  $m$ , of unstable modes to control). In this formalism, the beam pickup signals are processed in  $m$  parallel narrow-band filter channels, each with specified gain and phase response to have net 90 degree loop phase response at the modal oscillation frequency, with gain sufficient for net damping. The motion of every bunch is applied to each channel, with the filter selecting the superposition of the motion at the modal processing frequency. The  $m$  channels are then recombined in a single kicker output

channel to impress the control signal on the beam. This implementation becomes impractical and cumbersome for systems with thousands of bunches, many external HOM frequencies, external resonators with time varying center frequencies, filling patterns which vary in machine operation, etc - all which require changing the modal filter frequencies and gain/phase specification.

Time-domain processing implementations take advantage of the natural sampling that occurs in the bunched circulating beam with a single pickup and kicker. The formalism of bunch-by-bunch processing treats each bunch system as an independent harmonic oscillator, with an arbitrary external driving term which acts to couple the bunch to the other bunches in the accelerator. This leads to M independent processing channels for M individual bunches.

In each parallel channel, the correction signal for bunch n is found only from information from past samples of that bunch. This past history (sequence of turn by turn error signals) is processed (filtered) to create an output correction signal to apply on the next subsequent turn. The sampling of the pickup system inherently implements filtering in discrete-time formalism. The most general forms of these filters are Finite Impulse Response and Infinite Impulse Response filters, of the form

$$K_{out}(i) = \sum_{m=1}^M A(m)U_{in}(i-m) \quad (1)$$

$$K_{out}(i) = \sum_{m=1}^M A(m)U_{in}(i-m) + \sum_{m=1}^M B(m)K_{out}(i-m) \quad (2)$$

where the length of the filter  $M$  and the coefficients of the Filter ( $A_k$  and  $B_k$ ) determine the filter frequency response (for the FIR case, the sequence of the coefficients is the impulse response of the filter).

The necessary functions in the processing channel include i)removal of any DC offset (from static orbit offset in pickups, synchronous phase position), ii) selective gain at the oscillation frequency, with enough passband to function over an allowed range of machine tunes and modal tune shifts from reactive HOM impedances but reject out of band frequencies, iii) adjustment of the net loop phase shift as applied to the bunch (nominally 90 degrees for resistive damping without frequency shift but can be arbitrary considering placement of pickups, kickers, cable plant delays,

and group delays in processing channels, power amplifiers, etc.), iv) saturation control, to limit the dynamic range on injection transients, other conditions to avoid power stage saturation

The filters are designed to have maximum bandpass gain at the oscillation frequency, with net loop phase shift of 90 degrees including time delays and physical locations of pickup and kicker components. Techniques for filter design include analytic methods as well as numeric optimization[3] [4]. While logically N independent channels (one per bunch), the filters may be efficiently implemented in fewer than N processing blocks.

The bunch sampling acts to transfer (mix) the RF frequency of the bunch motion (as observed in a pickup) to baseband, where the aliased tune is processed in the discrete-time filter, and the filter output is translated back to the RF frequency through the sampling of the beam in the kicker. The baseband transfer function of each filter is therefore replicated as upper and lower sidebands around every revolution harmonic across the full span of modes, as well as multiples of this span.

A bunch-by-bunch system with identical control filters on each bunch is also an all-mode frequency domain system, with identical modal filters around every revolution harmonic[5]. The two formalisms have the same signals present in the pickups and kickers.

**Stability and filter requirements, impedance-driven instability dynamics** To achieve net negative damping, the external damping provided by feedback must exceed the fastest growth rate from external HOM driven motion. In the frequency domain the growth rate for mode  $l$  can be computed via summation of external impedances[1], for the general case of N equally spaced bunches, the longitudinal growth rate  $\tau_l$  is given by

$$1/\tau_l = \frac{I_0 f_{rf} \alpha}{2(E/e)Q_s} \Re(Z_{|,l}^{eff}) - 1/\tau_{rad} \quad (3)$$

$$Z_{|,l}^{eff} = \sum_{p=-\infty}^{p=\infty} \frac{\omega_{p,l}}{\omega_{rf}} \exp(-\omega_{p,l}^2 \sigma_r^2) Z_{|}(\omega_{p,l}) \quad (4)$$

$$\omega_{p,l} = (pN + l + Q_s)\omega_{rev} \quad (5)$$

where the summation of external longitudinal impedance  $Z_{|}(\omega_{p,l})$ , the scaling with beam current define the threshold current (at which  $1/\tau_{growth} - 1/\tau_{rad} = 0$ ) ( $\alpha$  momentum compaction factor). For the transverse case[6], the effective impedance summation is over  $Z_{\perp}$  where

$\beta_{\perp}$  is the beta function at the location of the impedance

$$1/\tau_l = \frac{I_0 f_{rf}}{2(E/e)} \Re(\beta_{\perp} Z_{\perp,l}^{eff}) - 1/\tau_{rad} \quad (6)$$

$$\beta_{\perp} Z_{\perp,l}^{eff} = \beta_{\perp} \sum_{p=-\infty}^{p=\infty} \exp(-\omega_{p,l}^2 \sigma_r^2) Z_{\perp}(\omega_{p,l}) \quad (7)$$

$$\omega_{p,l} = (pN + l + Q_{\beta}) \omega_{rev} \quad (8)$$

The necessary gain of an external feedback system can then be found as a function of growth rate

$$1/\tau_l^{net} = 1/\tau_l - 1/\tau_{fb} \quad (9)$$

longitudinal case

$$1/\tau_{fb} = \frac{f_{rf}\alpha}{2(E/e)Q_s} G_{fb}(\omega_l) \quad (10)$$

transverse case

$$1/\tau_{fb} = \frac{1}{2T_0} \sqrt{(\beta_p/\beta_k)} G_{fb}(\omega_l) \quad (11)$$

Where the gain is expressed as the ratio of effective output kick on the beam (deflection angle or energy kick) divided by the input signal amplitude measured at the pickup location (transverse displacement or time-of-arrival coordinate in radians). This overall gain is the product of gains in the pickups/front end, processing block, and output power stage/kicker functions. The pickup and back-end gains depend on the choice of detection and operating frequency, and in the transverse case on  $\sqrt{(\beta_p/\beta_k)}$ .

The dynamics of beam instabilities driven by ion and electron cloud effects (two-stream instability) is often more complicated than impedance-driven motion. This general form of feedback processing is still appropriate to control Ecloud and ion instabilities, acting to damp oscillatory bunch motion from an external perturbation. The rapid dynamics of these instabilities may require faster damping rates (higher gains) and broader bandwidths in the control filter channel.

### Analog vs. digital processing implementations

The pickup signals can be processed, and the correction signals computed in either analog or digital processing technologies. Both IIR and FIR filters require time delays of M beam revolutions ( $M=$ filter length), and the delay elements must have sufficient bandwidth to keep the individual bunch signals time-resolved. Such delays may be implemented in coaxial cable, surface acoustic wave devices, superconducting delay lines, or

optical fiber systems with adequate  $\tau * B$  delay bandwidth products. The filter computations then require multiplication and addition terms to compute the FIR sum-of-products. Some very simple implementations might be restricted to fixed tap coefficients of  $\pm 1$  and such "two tap correlator" filters (subtracting of the previous turn from the present turn signal) are commonly used to remove DC offsets in the processing channel (in the frequency domain the filter has notches at harmonics of the revolution frequency).

As transverse oscillations are typically inherently undersampled (the beam integer transverse tune is not sensed in the processing system) the phase control of the processing filter is computed at the fractional (aliased) tune and adjusted via choice of the filter coefficients. Alternatively, it is possible to use two pickups, separated in betatron phase by ideally 90 degrees, and implement a quasi-quadrature phase rotator by adding the two signals with variable signed coefficients[4] [7]. This approach trades off the extra complexity of the second pickup and receiver against the number of required multiply/add operations in the filter. For systems operating near the 1/2 integer tune, this approach may require two kickers with  $\pi/2$  betatron phase separation for useful time-invariant gain.

Digital implementations are increasingly common, as high-speed A/D, D/A and flexible computation blocks have become available. The digital processing functions can be flexibly and efficiently implemented in both fixed-instruction signal processors, and increasingly in field-programmable gate array (FPGA) technology. In these technologies sampling rates of 500 MS/sec. have been achieved. In designing such a parallel-processing system for a system with harmonic number N, the parallelism of M internal processing channels and the requirement that samples from a particular bunch always be resident in the same processing channel requires attention[3]. The reprogrammable and flexible nature of the DSP implementation is especially useful for application of a single architecture to multiple installations.

As typical synchrotron frequencies are very low compared to revolution frequencies, longitudinal systems can be usefully downsampled (only process and update bunch information every K turns, K=downsampling factor). This technique reduces the aggregate MAC/s in the filter by

$1/K^2$ , but adds extra group delay in the processing filter which limits the maximum achievable system gain.

Pickups and kicker functions must have adequate bandwidth to span all coupled-bunch modes, which may be 250 MHz or more on all-mode systems, though the designer has flexibility in choosing operating frequencies. Button and stripline pickups are common[8]. Electronic processing of pickup signals can be performed at baseband or heterodyned from a higher frequency (potentially higher sensitivity) with attention to the phase stability and timing sensitivity of a receiver operating at high frequencies. Transverse receivers measure the product of bunch current and transverse displacement, channels typically process 4 orthogonal pickups. Both baseband and heterodyned processing are possible, with delta-sigma hybrids or difference processing to compute nominal X, Y and intensity signals for each bunch (similar to BPM processing needs). Slow feedback loops to remove static DC orbit offset are helpful in using the dynamic range of the processing channel for oscillatory signals[9]. Figure 2 shows a transverse heterodyned processing channel with use of comb generator and delta-sigma hybrid processing[10].

The longitudinal detection pickup signal is also proportional to current. A longitudinal error signal can be made from a sum of orthogonal pickups (making it insensitive to transverse position) and phase-detected against a stable RF system reference. For greater sensitivity, detection at a higher RF harmonic can be made using periodic "comb generator" techniques[11]. A transverse detector in a dispersive region can also provide a longitudinal error signal after betatron motion is filtered out. As for transverse processing, a slow feedback loop to keep the reference oscillator phase centered with respect to the nominal synchronous phase is helpful to maximize gain ( $\cos(\phi)$  detector response, use of dynamic range). The useful operating harmonic is limited by the presence of propagating modes in the vacuum chamber near the pickup and necessary operating range considering injection and filling pattern gap transients.

**Processing channel dynamic range requirements** In damped equilibrium operation, the controlled beam motion is damped to the receiver noise floor, and the required dynamic range is the system minimum, requiring only dynamic range

from the noise statistics of the receiver. In practical operation, the processing channel must have adequate dynamic range to allow excitation of the beam at a finite excitation amplitude (from either externally driven motion, injection transients, grow-damp excitation measurements, etc.) without saturation of the processing system or of the output power stage. For processing systems with analog receiver front ends, digital processing filter systems, and analog output power stages the choice of the quantizing size of the input A/D and output D/A must be matched to the noise levels in the analog functions. In operation the feedback systems operate with the noise floor of the front end, multiplied by the operating gain, driving the output stage to an equilibrium output power level. The number of bits of numeric processing is really only meaningful relative to these signal/noise ratios. High dynamic range digital systems with 12 or 14 bit A/D and D/A components, with associated numeric dynamic range DSP processing blocks may not have any useful purpose after the maximum output level in the power amplifier(s), divided by the operating gain, is compared to the noise floor in the front end receiver(s). As the power amplifier output power is expensive, there are economic constraints selecting the minimum necessary output power. In practice, at damped equilibrium high-gain systems may operate with the output stages running at 25% to 75% of full-scale from the input noise floor, so that 6-8 bit quantizing A/D and D/A functions, with 16 bit DSP computational structures, are typically numerically sufficient.

Necessary system gain is determined by the required damping (or the largest unstable impedance driving the system), constraining a minimum gain floor. If the gain is chosen too high, the input noise may easily saturate the output stage, and out of band signals, after passing through the control filter, may also saturate the output stage. There is also a maximum gain possible due to delay-bandwidth limits in the implemented processing[12] [13]. Hadron systems without significant radiation damping require care in choice of gains and noise levels to prevent emittance increases due to system noise floors at high gains[14]. It may be beneficial to use time-varying gain on selected injected bunches, rapidly damp injection at high gain, and after damping of injection transients return their gain to the minimum gain required to control unstable motion.

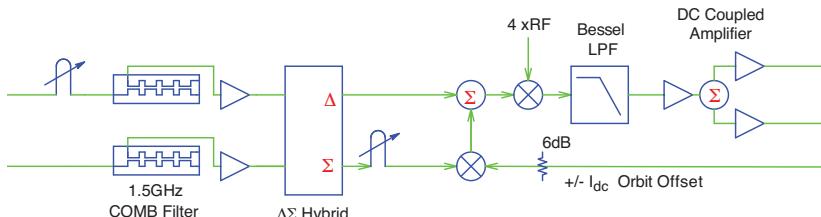


Figure 2: Front-end transverse receiver processing signals from opposed pickups with DC orbit offset via external adjustment.

The control filters reject DC from orbit offsets and synchronous phase position, though the necessary dynamic range of the A/D and DSP system can be reduced by the use of DC offset removal loops in the front-end. The current variation between bunches, and the synchronous phase transients in machines with high beam loading then define the minimum necessary dynamic range at the front-end ADC.

**Kicker and power stage implementation options** The necessary bandwidth and output power for the amplifier stage is determined by the minimum bunch spacing, the kicker design, bunch kick isolation requirement, and the highest frequency coupled-bunch mode which must be addressed. The power stages must have good phase linearity across the operating bandwidth for control of all modes.

The maximum output kick that is possible defines the maximum disturbed amplitude for which a linear controlled response is possible. For disturbed amplitudes below this level, the system returns towards equilibrium with exponential trajectories.

In operation, injection transients and other perturbations may excite a single bunch, or small number of bunches to a large level. Saturation in this case should be numeric, with saturated numeric levels chosen to be just below the analog saturation level in the power stages to prevent distortion from saturation recovery effects in a power stage from impacting signals applied to the remaining controlled bunches. Because the majority of the bunches stay in control, operation of these systems with a small number of saturated bunches is possible. The large amplitude bunches damp slowly due to radiation damping and the saturated kick, while the remainder of the beam is controlled near the nominal equilibrium point via correction signals.

The required bandwidth must span the lowest and highest frequency modes. For full power gain the risetime of the amplifier/kicker system to changes in drive level may require more than RF/2 bandwidth due to fill-time requirements in the kicker system (especially for travelling-wave striplines or drift tubes)

**Transverse kicker implementations** Most transverse kickers are transverse  $1/4\lambda$  stripline pairs driven differentially, with resistive terminations[8]. The response of such a transverse differential kicker has a *Sinc* frequency response. Systems have been developed which use open or short circuited transmission lines to kick countercirculating beams in a single power system [15]. For large circumference rings with very low frequency modes present (e.g. resistive wall instabilities) the necessary low frequency amplifier response may be difficult to achieve at baseband, and two band systems have been implemented [16] which use two separate kickers, with independent amplifiers, driving in two output bands to cover the required bandwidth.

Power levels of 50 - 200 watts per electrode are common, amplifier technologies are typically solid-state for systems with 125 - 250 MHz bandwidths. Lower bandwidth systems may use IOT or other tube technology drivers into capacitive transverse electrodes[14].

**Longitudinal kicker implementations** Longitudinal kickers can be implemented with gap structures [17], damped cavity kickers [18] or drift-tubes with delay lines[19]. For low frequency modes which can be controlled within the RF system bandwidth the RF cavity itself can be used as a longitudinal kicker. Transverse electrodes, driven common mode, can also make modest shunt impedance longitudinal kickers. Operating frequency choices depend on required

bandwidth and power. Operating frequencies in the 1 - 2 GHz band have been common for 250 MHz bandwidth. Amplifiers may be solid-state (paralleled GaAs or other semiconductor amplifiers) or TWT technology, though phase linearity and amount of AM-PM distortion are of importance to maintain system gain across the band and over operating power levels.

Multiple-amplifier implementations, with each amplifier driving each input port of electrode are common for high-power systems, with an advantage of greater operating robustness due to a failure of a single power stage, and disadvantage of  $n$  times as many precision timing adjustments to align the beam and feedback correction signal. Multiple kicker implementations are also common for high gain requirements. Operating the kicker systems at harmonics of the ring RF necessitates timing stability, means to adjust and control output signals with resolution of 10's of picoseconds.

**High-current considerations** High beam currents present challenges in power coupled to both transverse and longitudinal kickers. The designs must be free of spurious HOM impedances (often damped through coupling antennas or external waveguide terminations). Beam induced power in electrodes and feedthroughs for high current factory colliders can be  $>5\text{kW}$  per electrode or feedthrough, requiring great attention to cooling and monitoring of system connections and terminations. Cavity-type kickers have a cooling advantage compared to electrode or drift tube structures with internal vacuum electrodes, though the in-band directivity of the drift tube and transmission line structures does reduce beam induced power appearing at the power amplifier terminals. It is common to protect output stages from out of band power using absorptive filters in series with output power amplifiers and kicker drive inputs while absorbing in-band power in the amplifier. Circulators and isolators are also common, though these are relatively narrow band devices compared to the broad beam induced power spectrum, and through losses must be minimized to efficiently use the expensive output power. It is possible to offset the center frequency of a kicker system so that multiples of the RF system fundamental do not strongly couple to the structure (for example center frequencies of 9/4, 11/4 or 13/4 RF fundamental, in conjunction with single-sideband and QPSK modulation of the excitation signal to

flatten the frequency response across the operating band spanning all coupled-bunch modes[20]).

**Limits of feedback control** As error signals derived from pick-ups scale linearly with current, increases in operating current result in increased control gain. In principle this effect compensates the growth rates for impedance driven instabilities which also scale linearly with current. Naively stability is achieved without limit as currents increase, but the overall feedback loop gain is limited by stability limits due to group delay and required bandwidth. Due to limits of allowed feedback system gain, in conjunction with system front end noise and externally-driven motion of the beam causing saturation, achieving high current control requires a balance between impedance reducing techniques (HOM damping, design of vacuum structures, use of impedance-controlled RF systems with LLRF feedback) and instability control via feedback techniques. Practical limits in high-current colliders have been reached due to RF system noise[12].

#### 7.2.14.1 Beam diagnostics via feedback signals

Modulation of the system gain while recording beam motion (transient domain diagnostics) is a powerful beam diagnostic. Turning off feedback for a short interval, and recording the transient as beam instabilities grow, restoring the operating gain, and recording the damping transient provides a non-invasive measurement of impedance sampled by the beam. Transforms of the bunch data to modal domain reveal the free growth rates and tunes, and damping rates provided by the feedback [21]. Modal Frequency shifts between free and controlled states are a very sensitive measure of feedback system effective phase, and provide a means to carefully adjust the filter to have pure resistive damping[3]. Stable or slowly growing modes can be studied by driving a selected mode to finite amplitude before the transient recording. Measurement of the complex frequency of each mode allows frequency-resolved measurements of the HOM's driving the beam [22].

**Implementation examples** Systems have been implemented using minimal processing (single pickup, cable delay and power stage/kicker). An example analog transverse system from NSLS[23], Figure 3, uses betatron phase between

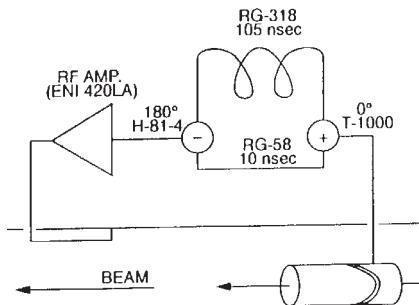


Figure 3: NSLS transverse feedback with 1 turn notch filter.

pickup and kicker, plus time delay, to approximate resistive damping (non-optimal phasing results in tune shift). Such a simple implementation has limited adjustability for operating tune, modest allowed gain, but is operationally useful.

More complex analog transverse system (Figure 4) as implemented at the ALS/Bessy-II uses two quadrature pickups, roughly 90 degrees apart in betatron phase, scaled and added for gain/phase adjustment with two-tap FIR delay line filters for orbit offset rejection[24] [7] [4].

Digital processing based systems have been constructed e.g. Fig.5 using fixed instruction DSP [25] [26], custom hardware [16] [27] and FPGA implementations [14] [28] [29] [30]. An example general purpose processor using reconfigurable FPGA processing, configurable for multiple rings, with common diagnostic and control software is shown in figure 4[4]. The use of common implementation platforms, common control and analysis software is very significant for economical and efficient development of new systems. The processing flexibility in the programmable systems is very significant for developing system diagnostics, beam diagnostics, and allows many beam measurements (e.g. bunch current monitoring) from the same hardware channel.

Table 1 shows 5 contemporary installations, representing control of transverse and longitudinal instabilities with achieved performance and some technology details. Examples are selected to show HOM, resistive wall and Ecloud interaction instability control in a contemporary light source, and the large factory colliders. In quantifying achieved damping performance, it is sometimes helpful to express damping/growth e-folding times in 1/turns (or oscillation cycles) to

compare installations with different revolution frequencies and oscillation frequencies.

## References

- [1] A. W. Chao, "Physics of collective beam instabilities in high-energy accelerators," New York, USA: Wiley (1993) 371 p.
- [2] M. Serio et al, EPAC96 p148
- [3] D. Teytelman, "Architectures and algorithms for control and diagnostics of coupled-bunch instabilities in circular accelerators," Ph.D. Thesis Stanford University June 2003 SLAC-0633
- [4] M. Lonza, "Multibunch Feedback Systems," Proceedings of CAS CERN Accelerator School on Beam Diagnostics:467-511,2009. CERN-2009-005
- [5] F. Pedersen, "Multibunch instabilities," Lect. Notes in Physics 425, Springer (1994) p269
- [6] J. M. Byrd and J. N. Corlett, PAC93 p3318
- [7] W. Barry et al PAC95 p2423
- [8] D. A. Goldberg, G. R. Lambertson, "Dynamic devices: A Primer on pickups and kickers," AIP Conf. Proc. 249, 537-600 (1992).
- [9] D. Boussard, Proc. 5th Advanced Accelerator Physics Course, CERN 95-06, vol. 1 (1995) 391
- [10] M. Tobiyama, KEK Laboratory
- [11] D. Briggs et al, PAC91, p1404
- [12] J. Fox, T. Mastorides, C. Rivetta, D. Van Winkle and D. Teytelman, PRST AB 13 052802 (2010) .
- [13] D. Teytelman , Invited Talk PAC11
- [14] W. Hofle, LHC-Lumi-06 Proceedings p.177
- [15] J. T. Rogers, et al , PAC95 p.2426
- [16] M. Tobiyama, E. Kikutani, PRST AB 3, 012801 (2000)
- [17] T. Kasuga, H. Yonehara, PAC89 p918
- [18] R. Boni et al, PA 52, 95-113 (1996)
- [19] J. N. Corlett et al, EPAC94, p1625
- [20] A. Gallo, M. Migliorati, L. Palumbo, NIM A404, 223-230 (1998)
- [21] S. Prabhakar, "New diagnostics and cures for coupled-bunch instabilities," Ph.D. Thesis Stanford University, SLAC-R-554, UMI-99-86494-MC, Feb 2000. 188pp.
- [22] D. Teytelman et al, PRST AB 4:112801,2001
- [23] D. M. Dykes, J. N. Galayda, PAC93, p582
- [24] S. Khan, T. Knuth, EPAC2000 p1912
- [25] D. Bulfone et al, PAC03, p3395
- [26] G. D. Stover et al, PAC99 p1213
- [27] J. T. Rogers et al, Fast digital transverse feedback system for bunch train operation in CESR, International Workshop on Collective Effects and Impedance for B Factories (CEIBA 95), Tsukuba, Japan, 12-17 Jun 1995 p. 475-480
- [28] E. Plouviez et al, , EPAC06 p2976

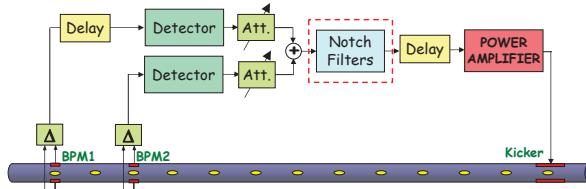


Figure 4: Two pickup quadrature front end, with notch filters to remove revolution harmonics (DC orbit offset).

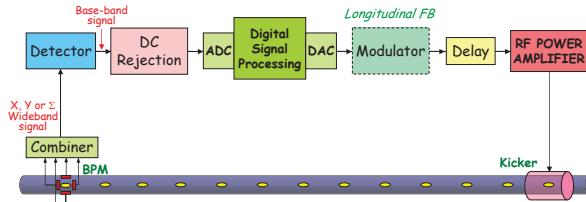


Figure 5: DSP based feedback system, showing general form of processing.

Table 1: Achieved performance for DSP feedback systems

|                     | Australian   | DAΦNE                       | KEKB                   | PEP-II         | PEP-II      |
|---------------------|--------------|-----------------------------|------------------------|----------------|-------------|
| SLS                 |              |                             | LER, HER               | HER            | LER         |
| I                   | 200 mA e-    | 1.1A e+                     | 1.7 e+                 | 2A e-          | 3A e+       |
|                     |              |                             | 1.3A e-                |                |             |
| plane               | V            | H                           | V,H                    | L              | L           |
| Instability         | HOM          | Ecloud                      | Res Wall,              | HOM            | Low cav     |
|                     |              |                             | Ecloud                 |                | mode        |
| Growth rate         | $2ms^{-1}$   | $26 ms^{-1}$                |                        | $0.38ms^{-1}$  | $3ms^{-1}$  |
| net damping rate    | $-15ms^{-1}$ | $-40ms^{-1}$                | -0.1 to -0.5 $ms^{-1}$ | $-0.75ms^{-1}$ | $-3ms^{-1}$ |
| e-fold (turns)      | 100          | 75                          | 200 to 1000            | 180            | 45          |
| e-fold (osc cycles) |              |                             |                        | 8              | 1           |
| notes               | Group Delay  | double system <sup>31</sup> | LER Ecloud             | RF noise       | Group Delay |
|                     | Limit        | $234ms^{-1}$                | dominates              | Limit          | Limit       |

- [29] T. Nakamura et al, EPAC04 p2649
- [30] iGp system, DIMTEL INC., San Jose CA.
- [31] A. Drago, PAC09, p3585

### 7.2.15 Beam Deflection and Collimation with Aligned Crystals

R.A. Carrigan, Jr., FNAL

Several useful books are available [1, 2]. Recent developments are covered in [3]. A formula collection can be found in [4]. There are web sites summarizing channeling work at IHEP and Fermilab [5].

**Critical angle** The limiting angle for capture to planar or axial channeling (related to Lindhard's critical angle) for a charged particle in a straight single crystal is

$$\psi_L = \sqrt{2U_0/pv} \quad (1)$$

where  $p$  is momentum and  $v$  velocity. For planar channeling  $U_0$  is the depth of the potential well and is proportional to  $Z_i Z N d_p a_{TF}$  where  $Z$  is the crystal atomic number,  $Z_i$  the beam particle atomic number,  $N$  the atomic number density,  $d_p$  the planar spacing [1], and  $a_{TF} \approx 0.8853 a_\infty Z^{-1/3}$ , the T-F screening parameter. At 1 TeV/c  $\psi_L = 6.7 \mu\text{rad}$  for the 110 plane in Si

where  $U_0 = 22.5\text{eV}$ . The effective critical angle is smaller because of atomic vibrations including thermal effects. Characteristically  $\psi_L$  is small compared to the typical beam divergence. Crystals must be aligned along planes or axes to see channeling. The critical angle is roughly equal to  $\psi_{1/2}$ , the HWHM of a transmission or alignment distribution of  $\psi$ , the particle angle relative to a crystal plane. However the distribution is not Gaussian. For a bent crystal

$$\psi_{Lb} = \psi_L(1 - R_T/R_m) \quad (2)$$

where  $R_m$  is the minimum radius of curvature and  $R_T$  is the Tsyganov radius (see below). Higher  $Z$  crystals, e.g. W, have larger  $\psi_L$  but large dislocation-free crystals are not available whereas they are for Si.

**Bent crystal channeling** As in a synchrotron, positive particles will stay in the channel of a gently bent crystal. At a small enough bending radius, the Tsyganov or critical radius [1],

$$R_T = p\beta c/eE_c \quad (3)$$

particles no longer channel. Here  $E_c$  is the interatomic field at a distance where the particle trajectory no longer remains stable (5.7 GeV/e [cm]) for Si(110) (Tab.1.3 of [1]). The equivalent magnetic field for relativistic channeling and a uniform bend with radius of curvature  $R$  is

$$B[T] = p[GeV/c]/0.3R[m] \quad (4)$$

Equivalent fields up to O(1000 T) are feasible. Practical problems that must be taken into account include bending crystals, unintended twists introduced in the process, nearby planes and axes, surface amorphous layers, and miscuts in a crystal face relative to a plane.

**Dechanneling** Particles dechannel because of multiple scattering in the channel and defects such as dislocations (Sec.2.5 of [1]). For planar dechanneling the dechanneling length for positive particles is

$$\lambda_d = \frac{256}{9\pi^2} \frac{p\beta c}{(\ln(2m_ec^2\gamma/I) - 1)} \frac{a_{TF}d_p}{Z_ir_e m_ec^2} \quad (5)$$

where  $I$  is the ionization potential (172 eV for Si). Subscript  $e$  refers to the electron.  $\lambda_d$  is an approximate exponential decay length. At 1 TeV, Eq.(5) gives  $\lambda_d = 51\text{ cm}$  for Si(110). The dechanneling lengths for negative particles are expected to be ten to twenty times shorter. In a bent crystal,  $\lambda_{db} = \lambda_d(1 - R_T/R_m)^2$ . At high energy, defect dechanneling should be dominated by dislocations. Nevertheless even at 7 TeV, the good Si that is available should still channel.

**Volume reflection and volume capture** In volume reflection [6] particles reflect off of planes when they are nearly parallel. The particles are deflected on the order of a critical angle away from the curving crystal. The process occurs over the whole arc of the bend. For collimation it can be cumulative for many passes. Since the expected deflection is  $O(\psi_L)$  the deflection will go as  $1/(p\beta)^{1/2}$ .  $R_m$  does not appear but  $\psi_L$  for a bent crystal will be a function of  $R_m$ . The efficiency of volume reflection can exceed 95% [7].

Volume capture (“feeding in” for a bent crystal) is a process where particles outside a channel in a bent crystal diffuse into the channel. Volume capture, first investigated by Sumbaev, et al. [2], deflects in the direction of the bend. This process is an analog of dechanneling where the particles instead diffuse in to the channel. As a result volume capture drops as the energy increases. The process occurs over the whole arc of the bend. Deflections can range up to the total bend.

**Bending transmission for external beams** In an external beam where emittance  $\gg$  crystal acceptance, the transmission efficiency is

$$F \approx F_s \left( \frac{\phi_b^{50}}{\Phi} \right) \left( 1 - \frac{R_T}{R_m} \right) e^{-s/\lambda_{db}} \quad (6)$$

$\phi_b^{50}$  = bent crystal phase space acceptance (proportional to  $\psi_{Lb}$ ),  $\Phi = 50\%$  beam emittance,  $F_s$  = crystal surface acceptance and  $s$  = crystal length. Note that  $\psi_{Lb}$  is for a bent crystal. For a harmonic potential in a straight crystal

$$F_s = \pi x_c / 2d_p \quad (7)$$

where  $x_c$  is the effective channel half width which depends on the bend, screening and lattice vibrations. Applications have included a beam throttle at Fermilab in NE at 800 GeV, a beam splitter to produce the  $K_s$  beam for CERN NA48 and a deflection of 150 mrad at Serpukhov.  $F$  was higher than 80% for a 400 GeV/c measurement at CERN [8]. The longest bent crystal has been a 15 cm exemplar at Serpukhov. Negative particles can be deflected but there is more dechanneling and  $\psi_L$  is smaller.

**Extraction** For an accelerator multi-pass extraction can occur where particles pass through a crystal several times before they channel [1]. The overall extraction and bending efficiency is

$$F_a = F_{ex} \left( 1 - \frac{R_T}{R_m} \right) e^{-(s/\lambda_{db})} \quad (8)$$

where  $F_{ex}$  is the extraction efficiency. Biryukov [1] simulations at 900 GeV give values of  $F_{ex} = 0.7$  with  $s$  optimized at 1 cm. Fermilab measurements with  $s = 4$  cm were about 30%. CERN measured 15% at 120 GeV/c. IHEP has reported 85% efficiency with 70 GeV protons. An approximate ansatz is

$$F_{ex} = \frac{1}{[1 + (\Theta_m/\psi_{Lb})(s/L_n)]} \quad (9)$$

where  $\Theta_m$  is the fitted multi-pass scattering angle and  $L_n$  is the nuclear interaction length. (Biryukov's Monte Carlo for  $F_{ex}(s)$ , Si, was fitted with  $\Theta_m$  of 38  $\mu\text{rad}$  whereas he had calculated 30  $\mu\text{rad}$ ). A more thorough treatment is contained in [9]. At high energy this formula favors very short crystals. Major steps toward short crystal lengths have been the development of so-called quasi-mosaic and anticlastic bending geometries at IHEP and PNPI. For an anticlastic geometry a bend on the length of a crystal produces a curve in the narrow width. IHEP has tested bends as short as 0.3 mm. Typically the crystal is placed at the beam edge and halo is pumped out to the crystal via a kicker, beam-gas scattering, noise or beam-beam collisions. Results from RHIC and Fermilab show alignment effects over the entire bend of the crystal, probably due mainly to volume reflection. This development is important for alignment, particularly for collimation.

**Collimation** (Sec.3.3.9) One suggestion for two stage collimation at the SSC was to replace the primary scraper with a bent crystal that would produce a substantial deflection on to a secondary absorber [10]. This technique could be 5-10 times as effective as conventional two stage collimation and be significant for high luminosity colliders. IHEP experience has shown the beneficial effect of crystals when used as primary collimators. Two stage crystal collimation has been successfully demonstrated at the Tevatron [11] where a reduction by a factor of two of some backgrounds was observed. Studies are now underway at the SPS. The promise of short bending length crystals further increases the chance of success. Volume reflection provides a tool for rapid crystal alignment. A CERN/Ferrara/IHEP/PNPI team has pioneered the development of multiple arrays of crystals for use in collimation with volume reflection. Some have been used for collimation studies at the Tevatron.

**Focusing** Crystals can be used as focusing elements. At Serpukhov, a 70 GeV p beam was

focused to a 40  $\mu\text{m}$  line using a crystal prepared by PNPI with a 0.5 m focal length [1].

**Spin precession** The spin of a channeled particle moving in a bent crystal should precess about the effective magnetic field through an angle

$$\phi = \frac{1}{2}\gamma\theta(g - 2) \quad (10)$$

where  $\theta$  is the deflection angle (Sec.2.6.1). The high effective magnetic field in a bent crystal offers an interesting possibility for charm particle magnetic moment measurements. The small angular acceptance is a significant limitation. This process has been demonstrated with Sigma hyperons at Fermilab by a PNPI-Albany-Fermilab group.

**Exotic acceleration and cooling** Channeling has also been discussed for damping beam emittance, that is beam cooling [12] and exotic acceleration [13]. Both possibilities are extremely challenging.

**Channeling radiation** [14], [16] For electron-positron colliders an interesting possibility is channeling assisted positron production [15].

**Radiation damage** Little degradation of MeV channeling in Si was observed after an exposure or  $4 \times 10^{20}/\text{cm}^2$  in a 28 GeV p beam. Si in a Serpukhov 70 GeV beam continued to channel after an exposure of  $10^{19}/\text{cm}^2$ . At CERN the crystal deflection efficiency was reduced by 30% for an exposure of  $5 \times 10^{20}/\text{cm}^2$  in a 450 GeV proton beam. Tests at SLAC by a Lyon-Orsay-MPI-CERN-Stanford collaboration have shown the absence of damage from electrons at similar fluences for crystals intended for use in channeling assisted positron targets.

## References

- [1] V.M. Biryukov, V.I. Kotov, Yu.A. Chesnokov, Crystal Channeling and its Application at High Energy Accelerators, Springer (1997)
- [2] Relativistic Channeling, Eds. R. Carrigan, J. Ellison, Plenum (1987)
- [3] S. Dabagov, L. Palumbo, Charged and Neutral Particles Channeling Phenomena – Channeling 2008, World Scientific (2010)
- [4] N. Doble et al, NIMB 119 (1996) 181
- [5] <http://crystalbeam.narod.ru/>;  
<http://home.fnal.gov/~carrigan/Channeling/Channeling master.htm>
- [6] A.M. Taratin, W. Scandale, NIM. B262 (2007) 340

- [7] W. Scandale et al, PRL 98 (2007) 154801
- [8] W. Scandale et al, PLB 680 (2009) 129
- [9] V.M. Biryukov, EPAC98 p2091
- [10] M. Maslov, N. Mokhov, I. Yazynin, SSCL-484 (1991)
- [11] V. Shiltsev et al, IPAC10 p1243
- [12] Z. Huang, P. Chen, R. Ruth, PRL 74 (1995) 1759
- [13] R.A. Carrigan, Jr., 7th Workshop on Advanced Accelerator Concepts, AIP Proc. 398 (1996)
- [14] M.A. Kumakhov, R. Wedell, Radiation of Relativistic Light Particles During Interaction with Single Crystals, Spektrum Physics (1991)
- [15] X. Artru et al, NIMB119 (1996) 246
- [16] H. Wiedemann in Handbook of Accelerator Physics and Engineering, 1st edn, 3rd print (2006) 217

## 7.2.16 Septum Devices

*R. Keizer, CERN*

A septum is a boundary that separates or combines beams by providing a different deflecting field on either side of the boundary. Such a field may be electrostatic or magnetostatic or electrodynamic in nature.

Fig.1 shows a septum configuration. The deflection of the beam is dependent on which side of the septum it passes. In most cases part of the beam will be intercepted by the septum causing radiation heating.

Although the bending power inside the magnet may be made very homogeneous, the diverging stray-fields near the magnet-ends are sources of important perturbations, which render the bending power non-homogeneous. This problem may be solved by introducing appropriately shaped end-plates which linearize the field.

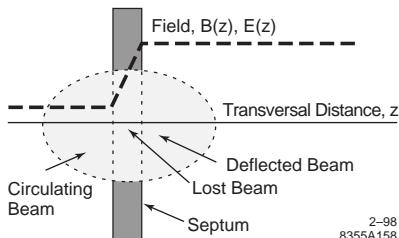


Figure 1: Thick extraction septum causing beam losses.

### Electrostatic septa (Sec.6.14)

In the case of electrostatic fields, the septum can be made very thin, <10 µm. Two possible

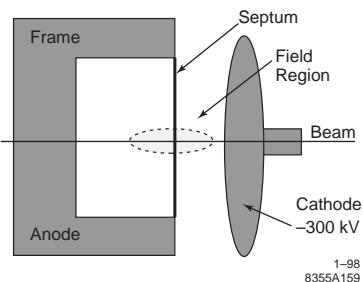


Figure 2: Electrostatic septum with external deflecting electrode.

configurations are discussed. The first one, Fig.2, employs a frame on which thin spring-loaded septum wires or flat metal strips are mounted. The opposite electrode, which can have either polarity, is mounted externally, allowing high extraction potentials >300 kV, but above-mentioned radiation heating will tend to deform the frame in such a way that the effective thickness of the septum increases. Because this is a run-away effect the frame should be made of Invar and the septum should be short.

In the second case, not shown, the electrode is placed inside the frame. The electric potential now will be considerably lower, typically 60 kV. The curvature, due to thermal deformation, has the same sense as the trajectory of the deflected particles, allowing thinner or longer septa.

The opposite electrode is usually made of metal such as stainless steel or titanium, causing high dark-currents or consists of anodised aluminium, which practically eliminates dark-currents and allows very high fields ~20 MV/m.

**Plane current-sheets** Imagine a 2-D cavity of arbitrary shape in the  $x, y$ -plane, Fig.3, the ferromagnetic walls run parallel to the  $z$ -axis. A number of arbitrarily positioned current conductors run also parallel to the  $z$ -axis and magnetise the cavity. The field density will be 2-D. At every point in the plane current-sheet,  $\vec{B}$  satisfies  $\nabla \times \vec{B} = \mu_0 \vec{J}$  where  $\mu_0$  is the absolute permeability,  $\vec{J}$  is the line current density. This means that  $\vec{B}$  vanishes everywhere in front of the current-sheet. The cavity can be made arbitrarily short by terminating it with a second current-sheet at the back using a coaxial current leadthrough. The basic principle of a septum may be summed-up by stating that the magnetic field should be terminated by its electrical analogue model.

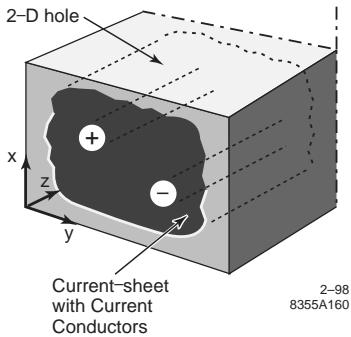


Figure 3: A 2-D magnetized cavity terminated by a plane current-sheet.

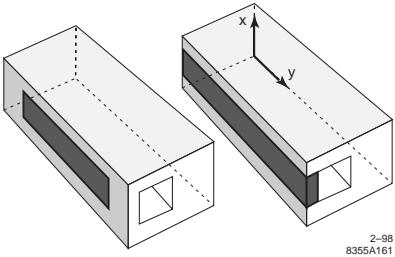


Figure 4: C-type septum dipoles with thick end-plates or extended septa.

*Application of the plane current-sheet theory to septum dipoles* In the case of an extended septum the basic cavity is of rectangular shape, Fig.4 (right), and the septum extends theoretically to infinity in both directions of the  $y$ -axis. For practical purposes it is necessary to reduce the length. The length,  $l_{\text{ext}}$  is chosen roughly one gapheight, see Fig.5, which makes the end field almost zero.

Multi-turn septa are also possible by cutting the current-sheets along the current flow lines as shown in Fig.5. Each conductor will have a different extension.

In the case of thick endplates, Fig.4 (left), the basic cavity is rectangular in shape but surrounded by ferromagnetic material on all sides. The end-plates have thickness  $l_{\text{ext}}$ . Since the flux is small, holes might be drilled to pass the beam.

**Transverse Field Calculations** The transversal fields in septum magnets can best be understood by evaluating the individual causes of field perturbations and then by summing them algebraically. This method allows easy evaluation of the influence of permeability, hysteresis, back-leg

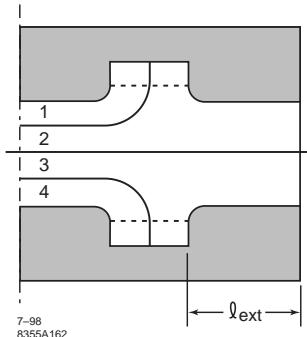


Figure 5: A horizontally split four-turn septum dipole.

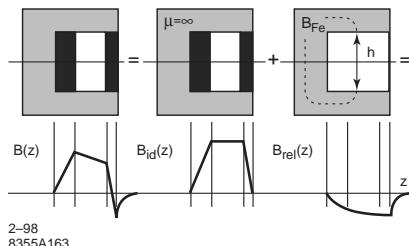


Figure 6: Field in a septum magnet with finite  $\mu$ .

windings or cooling ducts in septa. For finite  $\mu$  the analysis goes as follows. The field density  $B_{id}(x)$  in an ideal septum magnet with  $\mu = \infty$  will be

$$B_{id}(z) = \mu_0/h \quad (1)$$

inside and will be zero outside the magnet as shown in Fig.6. With finite relative permeability  $\mu_{Fe}(z)$  the field density in gap with height  $h$  is somewhat less by the amount  $B_{rel}(z)$ , where

$$B_{rel}(z) = h_{Fe}B_{Fe}/h\mu_{Fe}(z) \quad (2)$$

The length of the flux path is  $h_{Fe}(z)$  and  $B_{Fe}$  is the induction in the iron. The shape of the vertical field in an imperfect magnet may be obtained by summing the two functions as depicted graphically in Fig.6.

**Longitudinal field calculations** In the case of extended septa, where the fringe-field is almost eliminated, the dipole bending strength is constant everywhere in the aperture.

**Septum magnet construction** DC or slow pulsed septa require thick conductors due to Joule heating. Cooling is usually done by means of interior ducts or edge cooling. In the latter nonuniform temperatures lead to nonuniform  $J$  and thus

to possible stray fields.  $J$  as high as 400 A/mm<sup>2</sup> have been used for DC septa under special circumstances (see Sec.7.2.1)

Cooling problems are greatly eased by fast pulsing, e.g. half sine pulses (Sec.7.2.7) field distortions.

Apart from dipoles, septum multipoles, eddy current septa and steel septa (Lambertson magnets [9]) have been developed.

(Refs.[1]-[9] for septum magnets. Refs.[10]-[18] for pulsed magnets.)

## References

- [1] R.L. Keizer, CERN 74-13 (1974), contains a list of references between 1969 and 1974.
- [2] Y. Baconnier et al, Septum magnets for the extraction channels of the 400 GeV SPS (1977)
- [3] L. Evans et al, CERN-SPS/ABT/77-13 (1977)
- [4] A. IJsspeert, CERN SPS/ABT/AI/Int. 77-3 (1977)
- [5] R.L. Keizer, M. Mottier, CERN SPS/EA/78-16 (1978)
- [6] J. Dupin, CERN SPS/85-40 (ABT) (1985)
- [7] J. Dupin, A. Rizzo, CERN SPS/ABT/Note Technique 86-14 (1986)
- [8] S. Oliver, SPS/APS/Tech. Note 87-1 (1987)
- [9] G.E. Fischer, AIP Proc. 153 (1987) p.1190
- [10] L. Bartelson, J. Walton, FNAL-TM-1319 (1985)
- [11] P. Pearce et al, CERN/PS/BT/ 85-25 (1985)
- [12] J.A. Satti, S.D. Holmes, PAC 85, p3628
- [13] C. Ducastel, J.P. Royer, F. Voelker, CERN/PS/88-26(PO) (1988)
- [14] F. Neubau, DESY report HERA 90/05 (1990)
- [15] F. Rohner, CERN Note/AT/MA (1990)
- [16] L.R. Turner et al, PAC93, p2817
- [17] J.P. Royer, CERN/PS 95-13 (PO) (1995)
- [18] J.P. Perrine, M. Thivent, F. Voelker, CERN/PS 96-20 (PO) (1996)

### 7.2.17 Electron Lenses

*V. Shiltsev, FNAL*

Electron lenses employ EM fields of strongly magnetized high intensity electron beams and were originally proposed and built for compensation of beam-beam effects in hadron colliders [1, 2]. The lens employs a low energy beam of electrons which collides with the high-energy bunches over an extended length. Electron space charge forces are linear at distances smaller than the characteristic beam radius  $r < a_e$  but scale as  $1/r$  for  $r > a_e$ . Correspondingly, such a lens can be used for linear long-range beam-beam and

nonlinear head-on beam-beam force compensation depending on the beam-size ratio  $a_e/\sigma$  and the current-density distribution  $j_e(r)$ . Electron lenses have also been proposed for compensation of space-charge forces in high intensity hadron accelerators [3] (see Sec.4.14). Main advantages of the e-lenses are: a) the electron beam acts on high-energy beams only through EM forces, with no nuclear interactions; b) fresh electrons interact with the high-energy particles each turn, leaving no possibility for coherent instabilities; c) the electron current profile (and, thus, the EM field profiles) can easily be changed for different applications – see Fig.1; d) the electron-beam current can be quickly varied, e.g. on a time scale of bunch spacing in accelerators.

Two electron lenses were built and installed in the Tevatron p-pbar collider [4] and used for compensation of the long-range beam-beam effects [2], for the abort gap beam removal [5] and for halo collimation [6]. Upto 3 A, 6-10 kV e-beam was generated at the 10 mm diameter thermocathode immersed in 0.3T longitudinal magnetic field and aligned onto (anti)proton beam orbit over about 2 m length inside 6T SC solenoid. Electron pulses of 60-600 ns can be generated by changing electron gun anode voltage at the 48 kHz repetition rate.

## References

- [1] E. Tsyganov, A. Taratin, A. Zinchenko, Phys. Part. Nuclei 27 (1996) 279
- [2] V. Shiltsev et al, PRL 99 (2007) 244801
- [3] A. Burov, G. Foster, V. Shiltsev, FNAL-TM-2125 (2000)
- [4] V. Shiltsev et al, PRST AB 11 (2008) 103501
- [5] X.-L. Zhang, et al, PRST AB 11 (2008) 051002
- [6] G. Stancari et al, PRL 107 (2011) 084802

### 7.2.18 Spin Manipulation

#### 7.2.18.1 Siberian snake

*V. Ptitsyn, BNL*

Siberian snake avoids spin depolarizing resonances (2.6.3, 2.6.4). A full Siberian snake comprises one or several magnets (section 2.6.3)

**Solenoidal snake** The required field for 180° spin rotation around the longitudinal axis is:

$$HL[T - m] = 10.47 * p[\text{GeV}/c] \text{ electrons}$$

$$HL[T - m] = 3.75 * p[\text{GeV}/c] \text{ protons}$$

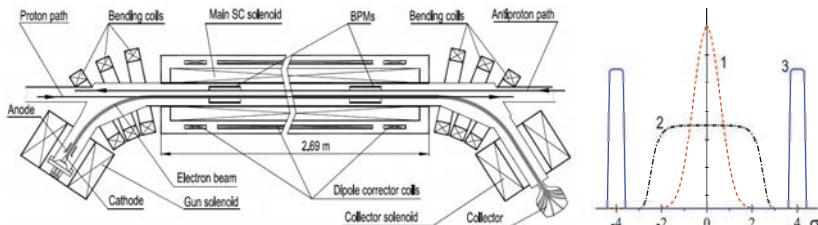


Figure 1: General layout of the TELs; (right) transverse electron current profiles for (1) space-charge and head-on beam-beam compensation, (2) for bunch-by-bunch tune spread compensation; (3) halo collimation.

The beam closed orbit is not affected, but the betatron coupling, introduced by the solenoids, may require a compensation. It is good for full snake application at low beam energies ( $< 5$  GeV). A superconducting solenoid snake was used for the proof-of-principle experiments for the Siberian Snake concept at IUCF (see Sec.2.6.5).

**Dipole magnet snake** The continuous axis snake was suggested by K.Steffen and generalized by S.Y.Lee (2.7.1):  $(-H, -V, m^*H, 2^*V, -m^*H, -V, H)$  where  $H$  and  $V$  present correspondingly horizontal and vertical bending magnets, and  $m$  is an integer. Due to internal symmetry of the Snake, the axis is in the horizontal plane and the beam orbit is restored at the Snake exit. By proper choice of the field, any direction of the Snake axis in the horizontal plane can be realized. The required fields scales inversely proportional to particle velocity and, hence, do not depend on the beam energy for relativistic beams. The orbit excursion inside the Snake increases inversely proportionally to the particle energy and exceeds 5 cm at the energies below 40 GeV. The integrated field is generally in the range 15-35 Tm, depending on the orientation of the Snake axis.

**Helical magnet snake** A continuous axis snake can be created also using four full twist helical magnets [1]. The corresponding symmetry conditions are:  $B_1 = -B_4$  ;  $B_2 = -B_3$ ;  $R_1 = R_4$  ;  $R_2 = R_3$ ;  $N_1 = N_4$  ;  $N_2 = N_3$  where  $B_i$ ,  $R_i$ ,  $N_i$  are, respectively, the field, helicity (+-1) and the number of the helical periods for  $i$ th helical magnet. The fields have to be vertically oriented at the entrance of each helix. The design used at RHIC is shown in Fig.1. Similar to a dipole magnet design, the magnetic field does not depend on the energy for relativistic particle. The resulting orbit excursion inside the Snake is considerably less than in the dipole magnet design (few cm down to 20 GeV). Since helical magnets have intrinsically nonlinear

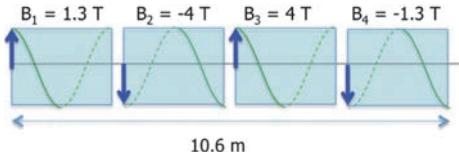


Figure 1: Helical Siberian Snake at RHIC. The choice of the fields corresponds to the Snake axis, oriented at  $45^\circ$  from the longitudinal. All magnets have the same helicity of the helical filed.

fields the effect on particle dynamics should be carefully evaluated.

### 7.2.18.2 Partial snake

*H. Huang, BNL*

A partial snake is a snake with rotating angle less than  $180^\circ$ . Similar to full snake, it can be built with solenoids, sequential dipoles and helical dipoles. With a partial snake in a synchrotron, a spin tune gap is generated when  $G\gamma$  is close to an integer  $n$ , where the spin tune  $\nu_{sp}$  is shifted away from the integer by  $\pm s/2$ . A partial snake is useful for medium energy synchrotrons such as the AGS and JPARC, as a full snake is not practical due to the large integrated field strength required and lack of long straight sections. Partial snakes can be used to overcome imperfection resonances only or to overcome both imperfection and intrinsic resonances. The strength for the first purpose is under 10%, and the strength for the second purpose is typically between 15% and 30%. The stronger partial snake is required due to spin tune spread, high order snake resonances for the later case. The helical dipole can be constructed with variable pitch such that the two ends have a helical pitch that is twice the pitch at the center. This field profile allows for a compact matching of the outside orbit to the helical orbit inside the

magnet. Both solenoidal [2] and helical [3] partial snakes have been used in the AGS. When they are used only to overcome imperfection resonances, an RF dipole is used to overcome the intrinsic resonances (see Sec.2.6.4).

For a strong partial snake, however, polarization loss due to spin mismatch at injection and extraction is no longer negligible. A 20% snake will lead to a 10% polarization loss due to this spin direction mismatch. This could be solved with appropriate spin rotators in the injection and extraction beam lines. However, a single additional partial snake located in the synchrotron can provide the spin direction matching at injection and extraction and also increase the effective partial snake strength if its position is chosen properly. The location and the precession axis direction of multiple partial snakes have to be chosen carefully to maintain control of the spin tune in a similar way as for multiple full snakes. Practically, the partial snake precession axis is always very close to longitudinal, which leaves only the location and strength of the partial snakes as free parameters. As an example, two equal partial snakes in the AGS can be separated by one third of the ring since it will introduce a periodicity of three units in the spin tune dependence on  $G\gamma$ . At the injection and extraction energies, for which  $G\gamma = 3n + 1.5$ , the two partial snakes cancel if they are equal strength. The polarization direction in the AGS is therefore exactly vertical and no polarization is lost due to spin direction mismatch in this case [4].

### 7.2.18.3 Spin rotator

*V. Ptitsyn, BNL*

Spin rotators are used to create a specific polarization orientation (most often longitudinal) at a specific location (for instance, in an experimental detector). In accelerator rings a pair of the spin rotators is usually installed, where the second rotator restores the polarization direction back to the vertical direction after the detector. Design features are similar to those of the snake designs described in Sec.2.6.3. Examples of spin rotator designs include rotators based on solenoids, dipoles and helical magnets. A solenoid rotates from vertical to horizontal followed by a horizontal bend to rotate into the longitudinal. Dipole magnets can be used to create longitudinal polarization at a fixed beam energy. A specific example is a “mini-rotator” comprising interleaved

horizontal (H) and vertical (V) bends. This was used in the HERA experimental regions to rotate the vertical electron spin (at 28 GeV) into the longitudinal [5] - (V,H,2\*V,2\*H,V,H) and vice-versa. The rotation by H&V is close to 45°. The total bend is not zero. For different beam energies or to flip the longitudinal spin orientation the transverse magnet position has to be changed to accommodate the change in orbit excursion. Helical magnets can also be used. The RHIC spin rotators are based on the sequence of four full-twist helical magnets (2.7.3): (B1+, B2-, B2+, B1-) where B1, B2 are the strengths of the magnetic field and the signs show the helicity. The magnetic field is oriented horizontally at the magnet entrances. This rotator is able to transform the vertical spin of polarized protons in any direction in the horizontal plane.

One full Siberian snake creates a stable polarization orientation coinciding with the snake rotation axis at the opposite azimuth of the accelerator ring, independently on the particle energy. Siberian snakes based on solenoidal magnets were used in AmPS (NIKHEF, Amsterdam) and Bates SHR (MIT, Boston) for this purpose to create the longitudinal polarization at the experiment locations. The snakes include two superconducting solenoids and a combination of the normal and skew quadrupoles to compensate for the betatron coupling [6].

A specific feature of the snake and spin rotator application in electron rings is that solenoidal magnets and vertical bends used in the Siberian snake and spin rotator schemes give rise to the dispersion  $\frac{dn}{d\delta p/p}$  of the stable spin direction, which defines the radiative depolarization (Sec.2.6.6). In order to minimize the  $\frac{dn}{d\delta p/p}$  function, or limit it only to the location of the spin rotator, so-called spin matching may be required (see Sec.2.6.8). The optics modifications to provide spin matching are strongly dependent on the spin rotator or snake scheme.

A comprehensive review of the Siberian snakes and spin rotator designs and the features of their application in the particle accelerators can be found in [7].

### 7.2.18.4 RF spin rotator

*M. Bai, T. Roser, BNL*

An RF spin rotator, a device which provides a time varying spin on the spin vector, can be used to excite a coherent spin precession by operating

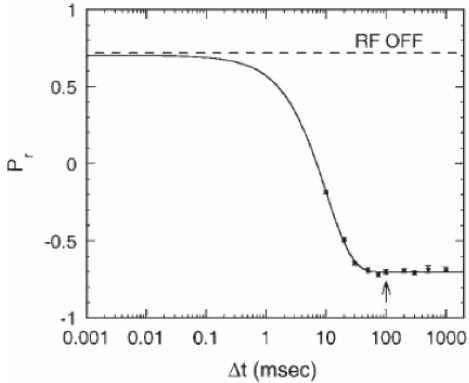


Figure 2: Spin flip using a single RF dipole at the IUCF. The dashed line is the beam polarization without RF dipole.

the RF spin rotator tune  $Q_{osc}$ , the frequency of the RF spin rotator in units of the particle's revolution frequency, close to the spin precession tune  $\nu_{sp}$ . This driven coherent spin precession can be used in spin manipulations such as inducing a full spin flip by sweeping the RF rotator frequency slowly across the beam spin precession frequency (Sec.2.6.4). By driving the RF rotator at a fixed frequency, one can also measure  $\nu_{sp}$  by measuring the amplitude of the projection of driven coherent spin precession in the horizontal plane [8]. A single RF dipole or solenoid, drives two artificial resonances at  $\nu_{sp} = Q_{osc}$  and  $\nu_{sp} = 1 - Q_{osc}$  (Sec. 2.6.4). When  $Q_{osc}$  is sufficiently different from  $\frac{1}{2}$ , contributions from the  $1 - Q_{osc}$  term oscillate at  $2Q_{osc}$  and adiabatically average to zero over many turns, and only one isolated resonance is excited. This artificial resonance was used to do precise beam energy measurement in an electron ring [9, 10, 11]. It can also be used in an accelerator where  $\nu_{sp}$  is not  $\frac{1}{2}$  to obtain a full spin flip by adiabatically ramping  $Q_{osc}$  through  $\nu_{sp}$ . This technique of using an RF solenoid was demonstrated at IUCF as well as COSY [12, 13]. Fig.2 shows the measured beam polarization as a function of the frequency sweep time of the single RF dipole in the IUCF [12].

A single RF dipole can also be used to enhance an existing intrinsic spin resonance by operating its frequency close to the vertical betatron frequency to excite a large amplitude coherent betatron oscillation. This can then result in a full spin flip and preserve the polarization.

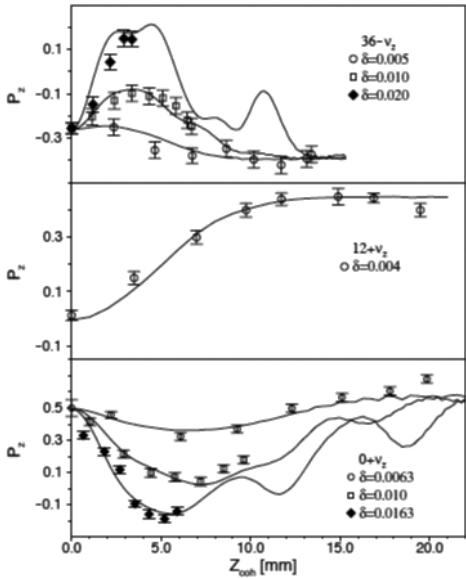


Figure 3: Beam polarization as a function of the size of the driven coherent oscillation induced by an RF dipole in the AGS at three strong intrinsic resonances in the AGS at BNL.

This technique was demonstrated in the BNL AGS [14]. Fig.3 shows the experimental results of using an RF dipole to overcome three strong intrinsic resonances at the AGS.

Another type of RF spin rotator is with rotating field, i.e. to rotate spin vector by a constant  $\phi_{osc}$  about an axis rotating in the horizontal plane at a frequency of  $Q_{osc}$  [15, 8]. The corresponding spinor equation of an RF spin rotator with rotating field is

$$\frac{d\psi(\theta)}{d\theta} = -\frac{i}{2}[G\gamma\sigma_3 + \phi_{osc}(\theta)e^{-i(Q_{osc}\theta+\chi)\sigma_3}\sigma_1]\psi(\theta) \quad (1)$$

where  $\phi_{osc}$  is the spin rotation amplitude of the RF spin rotator and  $\sigma_{1,2,3}$  are Pauli matrices. Eq.1 shows that only a single spin resonance is excited at  $\nu_{sp} = Q_{osc}$ . And the spinor transfer map  $M_{rot}$  is

$$M_{rot} = e^{-\frac{i}{2}\phi_{osc}(\cos(Q_{osc}\theta+\chi)\sigma_1 + \sin(Q_{osc}\theta+\chi)\sigma_2)}. \quad (2)$$

This can be implemented by placing a DC spin rotator between two RF dipoles. The two RF dipoles are operated at the same frequency with the same field oscillating amplitude and a fixed phase

difference  $\Delta\chi$  determined by the amount of spin precession from the DC spin rotators [8].

In reality, the driven coherent betatron oscillation due to an RF dipole forces particles to experience the fields from all the quadrupoles at the same frequency, and drives the spin resonance at  $\nu_{sp} = Q_{osc}$  as well as  $\nu_{sp} = 1 - Q_{osc}$  [16]. This can be mitigated by adding three additional RF dipoles to form two closed vertical orbital bumps to enclose the driven betatron oscillation within the spin flipper [8].

## References

- [1] V.I. Ptitsyn, Yu.M. NIM A 398 (1997) 126
- [2] H. Huang et al, PRL 73 (1994) 2982
- [3] H. Huang et al, PRL 99 (2007) 184501
- [4] F. Lin et al, PRST AB 10 (2007) 044001
- [5] J. Buon, K. Steffen, NIM A 245 (1986) 248
- [6] V.V. Danilov et al, Proc. 10th Intern. Symposium on High Energy Spin Physics, Nagoya, Japan, (1992) 445
- [7] S.R. Mane, Yu.M. Shatunov, K. Yokoya, J. Phys. G: Nucl. Part. Phys. 31 (2005) R151
- [8] M. Bai, T. Roser PRST AB 11 (2008) 091001
- [9] A.N. Skrinsky, Yu.M. Shatunov, Sov. Physics. Uspekhi 32 (1989) 548
- [10] L. Knudsen et al, PL B270(1991)
- [11] L. Arnaudon et al, PL B284(1992)
- [12] B.B. Blinov et al, Phys. Rev. Lett. 88(1) (2002) 014801-1
- [13] V.S. Morozov et al, PRST AB 7 (2004) 024002
- [14] M. Bai et al, PRL 80 (1998) 4673
- [15] T. Roser, Third Workshop on Siberian Snakes and Spin Rotators, BNL-52453, 1994.
- [16] S.R. Mane PRST AB 12 (2009) 099001

## 7.3 ACCELERATION

### 7.3.1 RF System Design

#### 7.3.1.1 RF system design for stability *D. Boussard, CERN*

##### Damping of phase oscillations [1]-[8]

###### Principle

(i) measure phase oscillations (PU electrode, filter at  $f_{rf}$  and phase detector); (ii)  $90^\circ$  phase shift of signal at  $f_s$ : differentiator, integrator (better if  $f_s$  changes much) or tuned filter (usually FIR in digital technology); (iii) apply correction to phase shifter in rf line (or on a Voltage-Controlled Oscillator, equivalent to phase shifter and integrator).

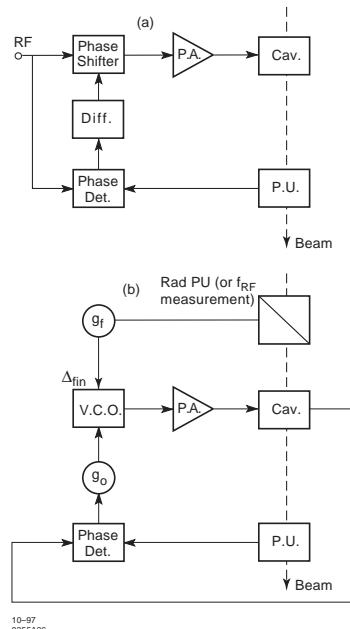


Figure 1: (a) Differentiator, (b) Phase lock loop with radial correction.

###### Practical differentiator

$$\text{Stability limit: } \tau_1 - \frac{1}{\omega_s^2 \tau_2} < \frac{1}{\omega_p} < 0 \quad (1)$$

$$\text{Low corner frequency: } \frac{1}{\tau_1} \gg \omega_s$$

$$\text{High corner frequency: } \frac{1}{\tau_2} \ll \omega_0$$

$$\text{Differentiator response: } s/\omega_p$$

$$\text{Beam response: } B = \omega_s^2 / (s^2 + \omega_s^2)$$

where  $s$  = Laplace variable,  $f_s = \omega_s/2\pi$  = synchrotron frequency,  $f_0 = \omega_0/2\pi$  = revolution frequency,  $1/\omega_p$  = differentiator gain.  
 Phase lock loop with radial correction

*Phase lock loop response* (pure integrator for  $G_0\omega_p$  large):

$$\frac{\Delta f_b}{\Delta f_{in}} \cong \frac{\omega_s^2}{G_0\omega_p} \frac{1}{s} \quad (2)$$

(See Fig.1b.)

*Closed loop response:*

$$\frac{\Delta f_b}{\Delta f_{in}} \cong \frac{1}{1 - s\tau_f}, \quad \tau_f = G_0\omega_p/G_f\omega_s^2 \quad (3)$$

$\tau_f \gg 2\pi/\omega_s$  for adiabaticity (condition maintained even close to transition where  $\omega_s \rightarrow 0$ ); VCO transfer function  $\omega_p/s$ .

**Rf noise [9]-[15]**Emittance growth rate (Sec.4.7.6)

$$\left\langle \frac{\Delta x}{\Delta t} \right\rangle = \frac{\omega_s^2}{4} [S_\phi(\omega_s) + 2x S_a(\omega_s)] \quad (4)$$

$$x = \sin^2 \frac{\hat{\phi}}{2}; \quad 0 < x < 1 \quad (5)$$

$\hat{\phi}$  is amplitude of phase oscillation.  $\langle \rangle$  means ensemble average over all possible noise having spectral density  $S_\phi$  (phase noise) and  $S_a$  (amplitude noise). Taking into account the synchrotron frequency spread the average over noise samples can be replaced by an average over particles, over a long time.

Usually phase noise dominates.  $S_\phi(\omega_s)$  may be strongly perturbed by the phase loop. If phase discriminator noise (spectral density  $S_u$  at  $\omega_s$ ) dominates (as it should normally) the emittance evolution is governed by the equation

$$\left\langle \frac{\Delta x}{\Delta t} \right\rangle \approx \frac{\omega_s^2}{4} \frac{S_u(\omega_s)}{16} x^2 \quad (6)$$

valid for short bunches with the solution ( $x_i$  = initial value),

$$x(t) = \frac{x_i}{1 - kx_i t}; \quad k = \frac{\omega_s^2}{4} \frac{S_u(\omega_s)}{16} \quad (7)$$

Equilibrium lifetime The applicable diffusion equation is

$$\frac{\partial \rho}{\partial t} = \frac{\partial}{\partial x} \left[ \frac{\omega_s^2}{4} (x S_\phi(x)) + x^2 S_a(x) \frac{\partial \rho}{\partial x} \right] \quad (8)$$

$\rho$  = particle density.  $\rho(x=1) = 0$  on the separatrix.

For amplitude noise (negligible loop effect) the equilibrium lifetime is given by

$$\tau_{eq} = 16 / [\omega_s^2 S_a(2\omega_s)] \quad (9)$$

Typical value of amplitude noise [15] is  $\langle \Delta a/a \rangle_{rms} \cong 5 \times 10^{-6} / \sqrt{Hz}$ . In the presence of phase loop, and phase detector noise dominant,

$$\tau_{eq} \cong 20 / [\omega_s^2 S_u(\omega_s)] \quad (10)$$

Measured phase detector noise at 200 MHz [15] is  $10^{-6} \text{ rad}/\sqrt{\text{Hz}}$  r.m.s.

**Robinson stability limits [1, 16]**

(see also Sec.2.4.9)

The beam-cavity system is stable (for short bunches) provided that two conditions are fulfilled ( $\gamma < \gamma_t$ ): "low intensity limit" is

$$\tan \phi_z > 0 \quad (11)$$

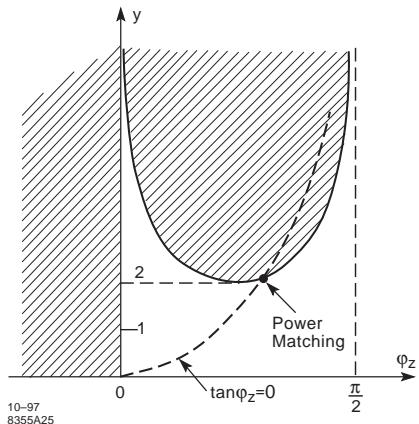


Figure 2: Robinson stability criteria. The system is just at the limit of stability for power matching, i.e. loading phase angle,  $\phi_L = 0$  (resistive load) and power delivered to beam =  $V^2/2R$ . (For a SC cavity matching means that all power from the rf source is transferred to the beam).

and "high intensity limit" is

$$Y \frac{\sin 2\phi_z}{\cos \phi_b} < 2 \quad (12)$$

where  $\phi_z$  = cavity impedance angle,  $Y = V_b/V$ ,  $\phi_b$  = stable phase angle (see Sec.2.4.3 for notation). The low intensity limit corresponds to a cavity resonant frequency which, for stability, must be slightly higher ( $\gamma < \gamma_t$ ) or slightly lower ( $\gamma > \gamma_t$ ) than the rf frequency. The high frequency limit corresponds to the center of gravity of the bunch sitting on the crest of the generator induced voltage: no stability for the dipole mode; exponential runaway of the beam phase. See Fig.2.

Locus of roots of characteristic equation

For  $i_b = 0$ , the characteristic equation has four roots, two purely imaginary at  $\pm i\omega_s$  and two purely real at  $-\sigma(1 \pm i \tan \phi_z)$ , ( $\sigma/2\pi$  = half cavity bandwidth). Increasing  $i_b$ , the two complex roots move, giving a complex frequency shift

$$\frac{\Delta \omega_s}{\omega_s} = \frac{\sigma^2 Y \tan \phi_z}{2 \cos \phi_b} \frac{1}{\omega_s^2 - 2i\sigma\omega_s - \sigma^2(1 + \tan^2 \phi_z)} \quad (13)$$

In the case  $\sigma \ll \omega_s$  and optimum detuning ( $\phi_L = 0$ , i.e.  $\tan \phi_z = Y \cos \phi_b$ ),

$$\frac{\Delta \omega_s}{\omega_s} \cong \frac{\sigma^2 Y^2}{2\omega_s^2} + i \frac{\sigma^3}{\omega_s^3} Y^2 \quad (14)$$

where the first term represents the positive frequency shift and the second, the Robinson damping term.

In case of  $\sigma \gg \omega_s$  and optimum detuning

$$\frac{\Delta\omega_s}{\omega_s} \approx -Y^2 \frac{\cos^2 \phi_z}{2} + i \frac{Y^2 \omega_s}{\sigma} \cos^4 \phi_z \quad (15)$$

where the first term represents the negative frequency shift and the second the Robinson damping.

One of the other two roots moves along the negative real axis with increasing  $i_b$  and reaches the origin at the high intensity Robinson limit.

**Stability of the  $n = 1$  mode [1, 17]** The critical parameter is the ratio of cavity detuning to revolution frequency,

$$\frac{\Delta\omega}{\omega_0} = \frac{1}{2} \frac{R}{Q} \frac{1}{V} h i_b \quad (16)$$

with  $h$  = harmonic number, and  $\Delta\omega/\omega_0 < 1$  for stability.

For stability: increase  $V$  per cavity (superconducting cavities); decrease  $R/Q$  (wide beam-tube aperture - superconducting cavities and, e.g. ARES cavities [17]).

### Stability with loops [1, 16]

With tuning loop only (simplified transfer function  $C_T = \omega_T/s; \phi_b = 0$ ),  $\omega_T$  = tuning loop gain, stability if

$$Y < \frac{\left[ 2(1 + \tan^2 \phi_z) + \frac{\omega_T}{\sigma} \right] \left( 2 - \frac{\sigma \omega_T}{\omega_s^2} \right)}{4 \tan \phi_z} \quad (17)$$

With phase loop only (simplified transfer function  $C_p = \omega_p/s; \phi_b = 0$ ),  $\omega_p$  = phase loop gain, stability for large loop gain if

$$Y < \frac{2}{\sin \phi_z}, \quad \phi_z > 0$$

$$Y < \frac{2}{|\tan 2\phi_z|}, \quad \phi_z < 0 \quad (18)$$

Phase loop extends the stability region only in the vicinity of the low intensity limit.

With tuning loop, phase loop and amplitude loop (simplified transfer function  $C_a = \omega_a/s$ ,  $\omega_a$  = amplitude loop gain), and  $\phi_b = 0$ ,  $\sigma$  large and transmission through beam unimportant,

$$Y < \sqrt{2 + \frac{\omega_a}{\omega_T} + \frac{\omega_T}{\omega_a} + \frac{\omega_p}{\omega_T} + \frac{\omega_T}{\omega_p} + \frac{\omega_a}{\omega_p} + \frac{\omega_p}{\omega_a}} \quad (19)$$

For a new design it is recommended that

$$Y \leq 1 \quad (20)$$

**Rf feed forward [18]** The beam signal is adjusted in amplitude and phase to generate via the power amplifier (PA) a current equal to  $-i_b$ . Static conditions are unchanged at the PA level (no additional power), but the system behaves dynamically as in the  $i_b = 0$  case, except for the tuning loop [1]. However, compensation is never perfect with 10% residual error being typical.

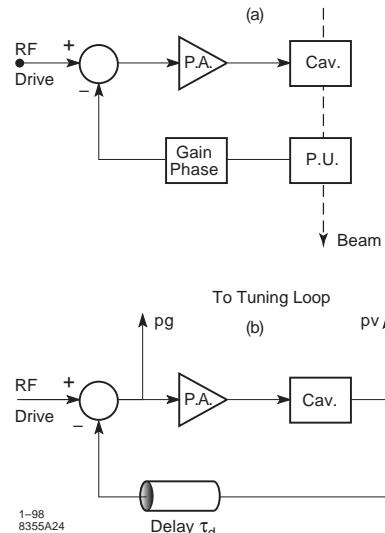


Figure 3: (a) Principle of rf feed forward, (b) rf feedback.

**Rf feedback [1, 6, 20]** Due to the unavoidable delay  $\tau_d$  (PA, cavity feeder, return cable), the open loop gain is limited for loop stability. For a simple RLC cavity, resonant frequency  $f_r$ , and, in the limit of large loop gain  $H \gg 1$ , the minimum achievable shunt resistance is

$$R_{\min} \cong 4 \frac{R}{Q} \tau_d f_r \quad (\pm 45^\circ \text{ phase margin})$$

$$\text{cavity half bandwidth } \delta f = 1/(8\tau_d) \quad (21)$$

Typical klystron delay is  $\sim 250$  ns; with special design it can be brought down to  $\sim 150$  ns.

Effect on other loops For outside loops (phase and amplitude) the cavity with rf feedback is equivalent to a cavity of much lower shunt resistance and the same  $R/Q$ .

Tuning loop response is modified by the presence of the rf feedback loop. For  $i_b = 0, \phi_L = 0$  the open loop gain is  $\propto 1/(1 + s/\sigma)$  (no rf feedback) and  $\propto 1/[1 + \frac{s}{\sigma(1+GR)}]$  with rf feedback

(open loop gain  $GR$  at  $f_r$ ). With  $N$  cavities driven by the same PA, the effect of rf feedback on tuning loop bandwidth decreases like  $1/N$ .

At  $i_b \neq 0$ , the sign of the open loop gain may be reversed. To keep a well balanced tuning loop (with rf feedback and  $i_b \neq 0$ ) replace the phase detector signal by a signal proportional to the normalized reactive generator power [19],

$$P_r = R(i_g/V) \sin(p_V - p_g) \quad (22)$$

$p_V, p_g$  are the rf phases of cavity voltage and rf generator.

Several cavities driven by the same PA [20]

The imbalance limits the minimum closed loop impedance  $Z_{CL}$  that can be achieved.

One cavity:

$$Z_{CL} = \frac{Z_C}{1 + GZ_C} \quad (23)$$

$Z_C$  = cavity impedance,  $GZ_C = H$ .

Two cavities:

$$Z_{CL} = \frac{2Z_C [1 + H\Delta G\Delta F(1 - \Delta Z^2)]}{1 + H(1 + \Delta G\Delta Z + \Delta G\Delta F + \Delta Z\Delta F)} \quad (24)$$

where  $\pm\Delta G$ ,  $\pm\Delta F$ ,  $\pm\Delta Z$  are the imbalances of generator coupling, feedback probe sensitivity and cavity impedance respectively. For  $H \gg 1$ , the symmetry tolerance becomes  $|\Delta G\Delta F| \ll 1/|H|$ . Typical values of  $\Delta G$  and  $\Delta F$  are  $\sim 10\%$ .

**Rf feedback with periodic filters** The open loop gain limitation given by the overall loop delay can be overcome if the total delay is made exactly equal to one turn ( $T_0$ ) and the frequency response of the return path of the loop is periodic with period  $f_0$ .

Simple comb filter (Fig.4a)

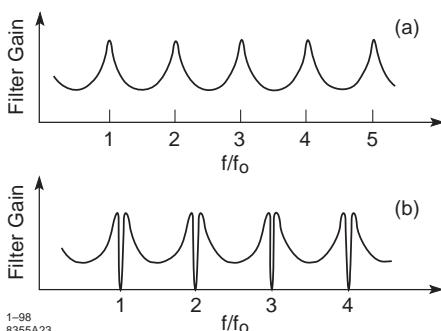


Figure 4: (a) Response of simple comb filter, (b) Response of double peaked comb filter.

Such a filter peaks at every harmonic of  $f_0$  and reduces  $Z_C$  in the vicinity of every  $f_0$  harmonic [6, 20, 21]. Overall filter and delay transfer function

$$\frac{G_0}{1 - ke^{-i\omega T_0}} e^{-i\omega T_0} \quad (25)$$

For one turn delay,  $0 < k < 1$ .

Double peaked comb filter (Fig.4b) [20, 22]

Such a filter peaks at every  $f_s$  sideband of every  $f_0$  harmonic and is better suited for constant  $f_s$ . The overall filter and delay transfer function is

$$\frac{G_0 (1 - e^{-i\omega T_0})}{[1 - ke^{-i(\omega T_0 + 2\pi\nu_s)}] [1 - ke^{-i(\omega T_0 - 2\pi\nu_s)}]} e^{-i\omega T_0} \quad (26)$$

$(0 < k < 1, \nu_s = f_s/f_0)$

## References

- [1] D. Boussard, CERN 91-04; CERN Acc. School 1991, CERN 92-03
- [2] S. Koscielniak, TRIUMF Design Note, TRI-DN-92-K216 (1992)
- [3] W. Schnell, Int. Conf. High Energy Acc. (1959) p.485
- [4] G. Schneider, CERN/PS 90-59 (RF) (1990)
- [5] F. Blas et al, EPAC94, p.1568; CERN/PS 94-24 (RF)
- [6] R. Garoby, CERN Acc. School (1996); CERN/PS 97-34 (RF)
- [7] L.K. Mestha et al, EPAC92 p1076
- [8] A. Gamp et al, EPAC90 p1551
- [9] S. Hansen et al, PAC77 p1452
- [10] D. Boussard, CERN Acc. School (1983) CERN 84-15
- [11] G. Dôme, idem, p.215
- [12] G. Dôme, CERN Acc. School (1985), CERN 87-03, p.370
- [13] S. Krinsky, J.M. Wang, PA 12 (1982) 107
- [14] J.A. Ellison, B.S. Newberger, H.J. Shih, PAC91 p216
- [15] D. Boussard, T.P.R. Linnecar, Proc. 12th Int. Conf. High Energy Acc. (1983) p.226
- [16] F. Pedersen, PAC75, p1906
- [17] T. Kageyama et al, EPAC96, p2008
- [18] D. Boussard, CERN SPS ARF Note 78-16
- [19] F. Pedersen, PAC85, p2138
- [20] F. Pedersen, CERN/PS 92-59 (RF)
- [21] D. Boussard, G. Lambert, PAC83, p2239
- [22] D. Boussard, E. Onillon, PAC93, p2379

### 7.3.1.2 Low level RF

*J.D. Fox, SLAC*

**Accelerator LLRF systems** Low Level RF systems provide control of frequency, amplitude and phase of cavity fields via signals and associated modulations driving high-power RF amplifiers and accelerating systems. LLRF systems synchronize multiple RF power sources and cavities to a common reference. They often incorporate feedback regulation to control accelerating cavity RF fields and reject external noise from RF power supplies and external disturbances. The RF cavity fundamental impedance, and its interaction with the particle beam, can be modified via the use of LLRF wideband feedback (direct loop) and periodic narrowband (comb loops). In many implementations beam longitudinal motion is stabilized via signals and modulations injected into the RF processing chain.

The general form of cavity RF voltage feedback senses a cavity RF signal, acting back through an error term through the power stages to regulate the cavity RF fields. If the reference is itself a time-varying signal, the topology forces the RF cavity fields to follow the reference in the manner of a servomechanism, such as required for frequency program in a cyclotron. External disturbances, such as beam injected signals, or external modulations on the high-power RF from power supplies, etc. are attenuated proportionally by the loop gain of the feedback path at the disturbance frequency. Any regulator with feedback has bandwidth and stability limits on loop gain and phase shift including group delay around the feedback path. For LLRF systems in accelerators, these limits may constrain regulation to finite operating bands around RF fundamental carriers or specific periodic frequencies where the feedback is active.

This regulator/servomechanism formalism is applicable to both pulsed RF systems as used in linacs and CW systems used in circular machines and CW linacs.

Implementation of the feedback path may be done directly at the RF frequency or (more commonly) heterodyned to an IF frequency or to baseband. Such approaches require complex processing, with RF signals represented in polar form (amplitude and phase) or as orthogonal real/imaginary signal components. Regulators may be constructed in analog or digital technologies, each with advantages/disadvantages of

flexibility, complexity, group delay (limits bandwidth and gain) and diagnostics.[1, 2]

A pulsed LINAC system topology is shown in Figure 5[3] with feedforward regulators modulating the RF drive to the power stage during the RF pulse. During the active pulse feedback control is possible within the loop bandwidth, in conjunction with feed-forward modulation of the reference within the pulse to shape or control the ultimate accelerating field applied to the beam during transit in a travelling wave structure. An error vector is generated on each pulse, and two low-bandwidth proportional-integral loops adjust the contents of the feedforward error vectors between pulses. This adaptive correction signal may contain high-frequency components, such as compensation for reflections and active phase/amplitude control to compensate for beam current structure, modulator ripple or other time-varying power requirements, though the updates and adaptation of the drive vector is band-limited to a fraction of the repetition rate. Over time, this topology adjusts to drifts and uses history of past pulse responses to adapt the drive signal to minimize the error in the high-power fields in the LINAC structure pulse (adaptation rate restricted by the sampling frequency at the pulse repetition rate). The example from the SNS achieves 50 kHz closed-loop bandwidth for the feedback loop, while the use of the adaptive feedforward algorithms compensate for beam loading during the pulse, Lorentz detuning of the superconducting cavities, etc. with 10 to 20 pulse time constants. The feedforward compensation for high beam loading conditions acts to increase the drive power to the klystron by x16 in conjunction with the beam arrival, avoiding saturation of the feedback path and improving transient response and loop stability. The combined overall regulation is better than 1% in cavity amplitude and +/-0.5 degree in phase [3]. Measurement of closed-loop regulation errors require an independent out-of-loop reference to quantify the impact of noise from internal sources acting within the feedback loop.

CW Linacs and circular machines typically use feedback regulators to control cavity fields, often the controllers are implemented on a per cavity or linac section basis, with overall global phasing between stations required to maintain the proper vector sum of the fields [4]. Such systems may be in analog or digital form. Recent digital

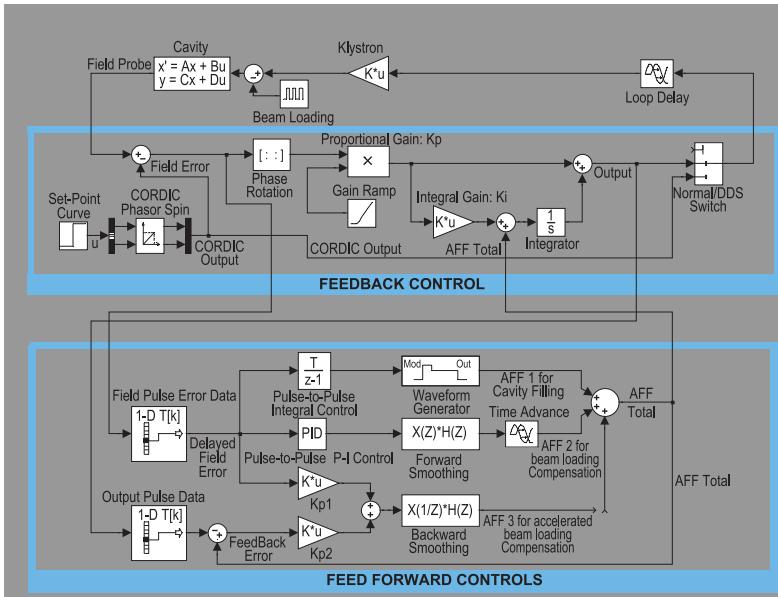


Figure 5: LINAC pulsed LLRF system from SNS.

control techniques as implemented for storage rings and for ERL have low-latency digital control loops. [1, 5]

Circular machines with high beam loading benefit from RF feedback techniques to regulate the RF accelerating fields (may also incorporate feedforward techniques to compensate for bunch gap transient effects [6]), as well as control low-mode coupled-bunch longitudinal instabilities driven by the fundamental impedance. An example system from LHC incorporates a wideband analog direct loop, as well as a digital comb loop (first order infinite impulse response filter) to implement cavity voltage regulation and impedance control. The digital polar loop acts to linearize the klystron complex frequency response which is nonlinear as a function of power level. Slow regulators for other functions, plus gap transient feedforward modulation are not shown in the figure [7].

The system Setpoint reference input sets the desired cavity RF magnitude and phase. The system acts to regulate the cavity field around this setpoint through the Analog path as well as the Digital integrator path. The beam current acts to change the cavity fields, the direct and comb loops feed back this disturbance through the

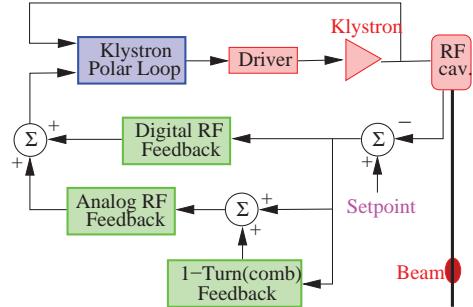


Figure 6: LHC LLRF block diagram.

power stage, reducing the effective voltage generated by the beam current. The effective impedance seen by the beam is reduced in two ways - a wideband reduction via the direct loop, plus comb path narrowband reductions at each synchrotron sideband around each revolution harmonic within the loop bandwidth.

Figure 7 shows in baseband representation the reduction of the naked cavity impedance, the impact of the direct loop in reducing the magnitude but spreading the impedance over many revolution harmonics, and the narrowband comb loops further reducing the impedance. The effective

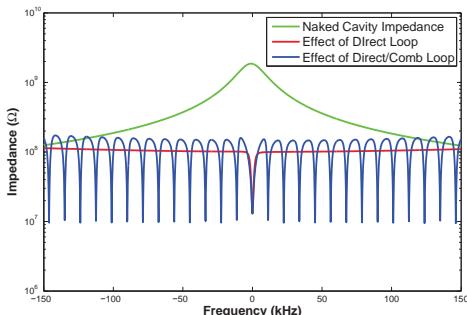


Figure 7: Effective LHC cavity impedance.

cavity impedance for mode zero is reduced by 26 dB (direct loop) and additional factor of 20 dB (comb loop) at the operating Q of 20,000. [7] The gain and bandwidth of the direct loop is limited by the total loop group delay, of note is the detail that the comb loop rotates in phase multiples of  $2\pi$  in this full 300 kHz closed loop bandwidth, but stability is retained as in the positive feedback phase regions the gain is  $< 1$ .

Excellent contemporary reference material is found in presentations at the LLRF workshop series (e.g. <http://www-conf.kek.jp/llrf09/llrfintro.html> and other years)

## References

- [1] Larry Doolittle, PAC09, p.1879
- [2] J. Delayen, SRF 2007, “LLRF Control Systems - Tuning Systems”  
[http://accelconf.web.cern.ch/AccelConf/srf2007/TUTORIAL/PDF/Tutorial\\_4.pdf](http://accelconf.web.cern.ch/AccelConf/srf2007/TUTORIAL/PDF/Tutorial_4.pdf)
- [3] M.A. Hengjie et al, LINAC06, p.571
- [4] H. Schwarz, M. Tigner, in Handbook of Accelerator Physics and Engineering, World Scientific (2006) p.552
- [5] M. Liepe et al, PAC05 p.2642
- [6] P. Corredoura, PAC99, p.435
- [7] P. Baudrenghien, EPAC06 p.1471

### 7.3.2 Klystron Amplifiers

#### 7.3.2.1 Klystrons

*D. Sprehn, SLAC*

The understanding and implementation of klystron design has not altered drastically in the past 50 years and is discussed well in the literature [1]–[6]. The Mark-III klystron (30 MW S-band 1-microsecond pulse) developed at Stanford 60

years ago is the predecessor of high-power pulsed klystrons today. Lower-frequency high-power CW klystrons have also been around for many decades. Currently CPI, Thales, Toshiba and L-3 are the major industrial suppliers of accelerator klystrons.

A klystron exchanges the kinetic energy of an electron beam for useful rf energy. A pulsed or DC voltage is applied across a thermionic cathode surface and anode to accelerate an electron beam. The cathode, made from a suitable material with appropriate work-function, is heated enough such that ample electrons are available for acceleration. With proper design the electron beam is accelerated and travels through the center of the anode and down a drift tube while focused by an externally-supplied magnetic field (e.g. solenoid or PPM) and eventually impacts the collector. Between the gun and collector are a series of cavities, the first having an rf signal applied to it which modulates the beam velocity. While travelling down the drift tube bunches form (density modulation) and the rf power on the beam grows. As the bunches travel they react with more cavities that further modify the velocities to alter gain, bandwidth and efficiency. The bunches relinquish their energy while traversing a final output structure, often a single cavity gap, and the rf power coupled to a transmission line to then be used for powering structures in the main accelerator. Ceramic at the gun and output transmission line isolate the tube vacuum from the outside world.

For accelerator applications the klystron amplifier is preferred over other devices due to the available output power at a given efficiency on a cost-per-watt basis. Typically a single klystron, operating in the range of  $\sim 1$  MW to  $\sim 100$  MW, Fig.1, will drive discrete cavities or multiple cells of standing-wave or travelling-wave structures via some rf distribution system. The choice of rf source is of course application-specific as some lower power devices (e.g. magnetrons, IOT's) have been used in accelerators and are proposed for future accelerators by simply using more of them.

Accelerator klystrons typically operate in the range of  $\sim 100$ -400 kV,  $\sim 0.4$ -12 GHz and higher, efficiencies of  $\sim 40$ -65%, gains of  $\sim 40$ -55 dB, bandwidth of  $\sim 1\%$ , peak output power of  $\sim 1$  MW-100 MW with average rf output power as high as  $\sim 1$  MW. The available efficiency is tied to the power output via the beam impedance

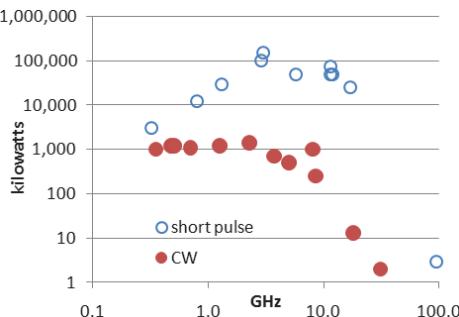


Figure 1: Power vs. frequency of modern klystrons. Note roll-off at lower frequencies is market-driven and not technology-limited.

and scales with the beam permeance  $K \equiv I_0/V_0^{3/2}$ , where  $I_0$  and  $V_0$  are the beam current and voltage. With high permeance a klystron achieves high power but lower efficiency due to space charge effects. An empirical formula for the peak efficiency using existing klystron designs is  $\sim 0.75 \times \mu K$  where  $x$  is a factor in the range of 0.15 to 0.2. For a balance of good efficiency and low voltage, klystrons generally are built with  $\mu K$  from  $\sim 0.5$  to  $\sim 2.0$ .

After selecting a voltage the klystron design considers factors such as modulator type, the ramifications of the voltage (physical sizes, flux densities), the overall system operation (gain, BW, pulse-length) and other frequency-related issues. This latter point is important as the beam cross-section is limited due to the size of the cavities and this affects beam formation (hence the cathode lifetime), the cost of focusing magnets, and the cavity coupling (which affects efficiency). Just because a reasonable voltage is selected does not mean that the efficiency or life-cycle cost is attainable due to these other constraints.

The beam drift tube radius “ $a$ ” is limited in size to ensure good coupling via the normalized radius,  $\gamma a = \omega/v_b^* \sqrt{(1-v_b^2/c^2)} a$ , where  $v_b$  is the beam velocity, and for good efficiency should be kept below unity. With  $b$ , the beam radius, kept at  $\sim 0.5a$  to  $0.75a$  the beam current density is known and the required cathode area convergence found. Convergence  $>\sim 100:1$  begin to become difficult (or costly) in a production environment and cathode current densities much over  $\sim 10\text{A}/\text{cm}^2$  for  $\sim \mu\text{s}$  pulses and  $\sim 2\text{A}/\text{cm}^2$  for CW tubes suffer from reduced lifetimes. Once  $b$  has been considered, the minimum focusing field required can

then be estimated from  $B_b = 832 I^{1/2}/(b V^{1/4})$  where  $B$ ,  $I$ ,  $b$  and  $V$  are in gauss, amps, cm and volts.

Bandwidth, normally not required  $>1\%$  for accelerator applications, can be increased to 10% or more by using more stagger-tuning of the cavities, coupled-cavities, cavity-loading or harmonic-cavities. These methods affect the gain, power or efficiency and the manufacture cost, operation or lifetime. Depending on the outcome of the previous design considerations, a compromise may be required.

Multiple-Beam Klystrons (MBK) (see also Cluster Klystron) operate at lower voltages than the single-beam klystrons of the previous discussion and therefore potentially increase the available design space. An MBK does this because it is in essence simply a grouping of low permeance (high efficiency) klystrons made in a single vacuum envelope (high power). When examined as a whole the permeance is high but the efficiency obeys the empirical formula for each beam individually. The main benefit for power tubes is one of lower voltage at a given power and efficiency which widens the choices of modulator and tank and possibly removes the necessity of oil insulation for very low voltage designs. This also may allow for wider bandwidth due to the lower beam impedance, and the larger surface areas can ease other power restrictions. The down-side of an MBK is in its cost, complexity and life-cycle details. When fielding relatively few devices it is likely that a standard klystron has significant cost advantages though the analysis of the relative merits of using MBK's is understandably complex. In addition to an MBK there are other beam transverse shapes (e.g. a flat sheet or a ring) which attempt to use larger surface areas and may lead to practical devices in the future. Typically, phase noise and stability issues are improved by stabilizing the beam voltage supply since the tube parameters of gain, frequency response and transit time are affected by voltage. Other affects such as the cathode heater supply, vibration and the rf drive may be mitigated by phase-correction circuitry operating at longer timescales.

**Practical usage** Proper care of a klystron should be practiced to minimize life cycle costs. It is important to interlock peak beam voltage and current, vacuum pressures, heater power, magnet power supplies, water flow and temperature, and odd behavior in the rf waveforms. It is also

worthwhile to perform cathode emission diagnostics, such as determining the work function [7] of the surface, at periodic intervals. Load reflection, potentially damaging to windows and the output coupling, may be mitigated at lower frequencies and power levels by ferrite-loaded isolators. At higher frequencies and power levels such devices do not exist but it is possible to use power-combining and/or delay time layouts to minimize the reflection. Arc detectors may be employed to examine transmission lines as further interlocks. With proper design, processing, interlocks and care, klystrons can achieve 100,000 hours MTBF.

## References

- [1] J.R. Pierce, Theory and Design of Electron Beams, Van Nostrand (1950)
- [2] T. Wessel-Berg, Space Charge Wave Theory of Interaction Gaps and Multi-Cavity Klystrons with Extended Fields (1960) Norwegian Defense Res. Inst. Bergen, Norway, Report No. 32
- [3] R.S. Symons, Scaling laws and power limits for klystrons, IEDM Digest (1986)
- [4] M.J. Smith, G. Phillips, Power Klystrons Today, Research Studies Press Ltd. (1995)
- [5] G. Caryotakis, SLAC Pub 10620 (2005)
- [6] G. Caryotakis, Handbook of Accelerator Physics and Engineering, 3rd Printing, , World Scientific (2006)pp. 549-552
- [7] M. Cattelino, G. Miram, Prediction cathode life expectancy and emission quality from PWFD measurements, Applied Surface Science Vol. 111 (1997)

### 7.3.2.2 Klystron amplifier systems

*A. Gamp, DESY  
M. Tigner, Cornell*

Klystron amplifiers are now in wide use for narrow band ( $\sim 1\%$ ) accelerator power amplifier applications above 350 MHz and above 10 kW peak or average power although IOT (see 7.3.4) and solid state amplifiers are making inroads at the lower power and frequency end of the spectrum. Three regimes in common use are: (i) short pulse ( $< 10 \mu s$ ),  $P_{pk}$  up to 100 MW; (ii) long pulse (0.1-1 ms),  $\bar{P}_{pk}$  to 10 MW; (iii) CW, power up to 1 MW. The generic schematic for all of these applications is shown in Fig.2.

**Short pulse** These are typically used for powering multicell linac structures with 10's MWs. Referring to Fig.2 (1) The master oscillator provided constant frequency during the filling time of

the accelerator structures. Typically the 100 kHz sidebands need to be  $-120$  dBc/Hz or better. Frequency distribution systems use phase stabilized coaxial cables [1] with the addition of phase feedback. More recently optical fiber systems [2, 3] provide an order of magnitude improvement in the achievable stability. With elaborated measures the optical technique can provide sub fs synchronization (see 7.8).

(2) The LLRF system provides general control of the amplitude and phase of the accelerating fields. (See the extended discussion in 7.3.1.2)

(3) Forward and reflected signal samples are usually derived via directional couplers from the input and output of the accelerating structure for monitoring the power flow and allowing diagnosis of breakdown phenomena or measurement of beam induced fields in the structure [4].

(4) The pulsed power to drive the klystron beam is derived from a dc power supply whose output is converted to high voltage, high current pulses by a line type modulator formed by a network of inductors and capacitors (see 7.2.7). These are tuned for a flat top pulse except for a ripple caused by the graininess of the network [5]. For pulses less than about 1 microsecond in length, transmission lines are used, providing a very flat pulse but are physically very long. The efficiency of pulse modulators of 70-75% is dominated by the unusable amount of power in the rise and fall times caused by leakage inductance in the pulse transformer. Lower turns ratios have lower leakage L but require higher voltage capability pulse forming networks and switching thyatrons.

(5) Transmission networks for mega-watt pulsed power levels use evacuated rectangular waveguides and at higher frequencies, say X-band and above, TE mode waveguides to keep insertion losses low. Pulse compression networks to increase peak power (see 6.7). Increase in usable power by x3 is common.

(6) Protective circuitry usually measures excessive reflected power and suppresses subsequent pulses to stop arcing in accelerator structures and waveguides.

**Long pulse** The emerging use is for  $\sim$ ms pulse length superconducting linacs for XFEL and linear collider (ILC) use [6, 7]. Systems under construction or development are typically 1 ms flat-top, 10 MW pk, 5 to 10 Hz rep rate.

(1) The master oscillator must have phase and amplitude noise levels commensurate with

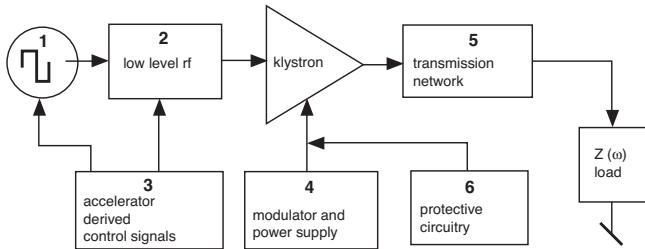


Figure 2: Typical klystron amplifier system with frequency dependent load impedance.

achieving well below 1 per mil zero current energy spread. The linacs, being km in dimension, require special synchronization means (see 7.7, 7.8)

(2) The LLRF has feedback on amplitude and phase with bandwidth 50kHz, (see 7.3.1.2 & Fig.3) The systems need to maintain bunch to bunch energy regulation to less than 0.1 per mil in the face of beam loading, beam loading fluctuations, Lorentz detuning of the cavities and microphonics. Overall amplitude and phase stability is such that synchronization on the 10 fs level could be demonstrated.

(3) Directional couplers and cavity pickup probes are used (see following sections)

(4) Three types of modulators are currently in play: i) bouncer circuit modulator [9] ; ii) pulse step modulator [10]; iii) Marx modulator [11]. i) and ii) utilize a pulse transformer for multiplying the voltage from the modulator to match the klystron requirement while iii) works by adding together a number of low voltage units. Type i) is the most thoroughly tested to date. Each is capable of the same parameters as shown in Table 1. Each can limit arc deposition to <20 J.

(5) As shown in Fig.3, [12] multiple cavities are powered by one klystron, the number depending on the cavity characteristics and operating gradient. Some attempt is made to match the peak gradient capability of each cavity with the setting of the input coupler and the overall voltage achieved by each group is regulated by the vector sum of the cavity voltages as shown[13]. Development has been done to make couplers that are easily adjustable [14].

(6) Same as short pulse above plus modulator arc energy limiting capability of <20 J.

**CW** These amplifier systems drive storage ring rf cavities and cw superconducting linacs now used for nuclear physics and coming into use as

radiation sources for materials science. (1) The MO's needed for the different applications are different. For storage ring service one expects something like 0.1° phase stability with the ability to be unlocked from the stable reference for chromaticity scans of a few x 100 kHz. Among linac uses the requirements differ depending on the exact application. Typical would be less than 100 fs jitter at frequency sidebands above 10 Hz. More stringent applications for pump-probe experiment timing will need sub 10 fs stability and will eventually need atomic clock stability (see Sec.7.8).

(2) LLRF for one current CW use [15] must deliver 0.5° in cavity phase and 4.4E-4 field regulation. Subsequent developments [16] have demonstrated 0.02° and 1E-4 respectively (see Sec.7.3.1.2).

(3) Directional couplers and cavity field probes provide the main information about the phase and field variables. Between 0.5 and 1 W at the cavity beamtube feedthrough is needed for sufficient signal in light of the attenuations encountered between 2 K and 300 K.

(4) For CW applications the dynamic heat load will be important and using as high a  $Q_{ext}$  as consistent with beam loading is needed to maximize efficiency in the face of microphonics where agile control of the individual cavity field is needed. This has lead to the choice of individual amplifiers for each cavity in CW applications [17] (see also Sec.7.1.1.8).

(5) Either waveguide or coax can be used depending on the amplifier-cavity distance. Temperature control of the transmission line will be essential.

(6) Fast cutoff for the amplifier on the microsecond scale is required to avoid window damage in the couplers, particularly in multi-kilowatt feeds.

## Feedback and Feed-forward applied to Superconducting Cavities

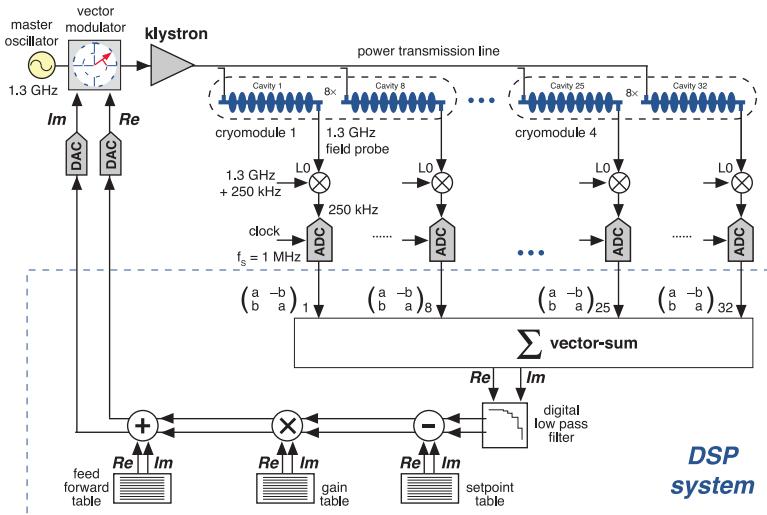


Figure 3: Schematic of XFEL distribution and LLRF system. One klystron is used to drive multiple cavities[12].

Table 1: [11] Characteristic load presented by the Multi-beam Klystron designed for this service.

|               |        |                |                  |
|---------------|--------|----------------|------------------|
| Peak Power    | 10 MW  | Beam Current   | 133 A            |
| Average Power | 150 kW | Pulse Duration | 1.5 ms           |
| Beam Voltage  | 115 kV | Efficiency     | >65% (saturated) |

### References

- [1] R.K. Jobe et al, PAC89, p1987
- [2] J. Urakawa et al, PAC91, p1555
- [3] A. Krycuk et al, PAC91, p1470
- [4] H. Schwarz, PAC85, p1847
- [5] A.R. Donaldson, 17th Power Modulator Symposium (1986), p. 230
- [6] M. Altarelli, R.Brinkmann et al, Ed's, XFEL-Techical Design Report ,DESY 2006-097. p.82
- [7] N. Phinney, N. Toge, N. Walker, ILC RDR, 2007, <http://www.linearcollider.org/about/Publications/Reference-Design-Report>
- [8] C.Schmidt, DESY Accelerators 2010, p.54, [http://www.desy.de/ueber\\_desy/jahresberichte/index\\_ger.html](http://www.desy.de/ueber_desy/jahresberichte/index_ger.html)
- [9] H.D. Schwarz, M. Tigner in Handbook of Accelerator Physics and Engineering (2006) p.553
- [10] H. Leich et al, LINAC10, p893
- [11] C. Burkhardt et al, IPAC10, p3636
- [12] T.Schilcher, Thesis, DESY and Universität Hamburg, (1998)

[13] S.Simrock et al, LINAC06, p.559

[14] C.Nantista, C. Adolphsen, LINAC06, p.643

[15] C.Hovater et al, LINAC10, p280

[16] M.Liepe et al, PAC05, p2642

[17] R.Nelson, A. Kimber, PAC11

### 7.3.3 Tetrode Amplifiers

*J.M. Brennan, BNL*

**Tetrode principles [1, 2, 3]** Tetrodes are used to power cavities up to 300 MHz and 100 kW. They are preferred for stability and high power gain and act approximately as a voltage controlled current source (transconductance device) which can be coupled to the cavity by some impedance matching device. Commonly driven class AB<sub>1</sub>, efficiencies in the range 60% are achieved. Transit time effects limit gain and very high frequencies.

Thoriated tungsten cathodes heated to  $\sim 2000$  K by heater currents of  $\sim 500$  A at 50/60 Hz are normally employed. Other relevant quantities are:

## Tetrode Characteristics

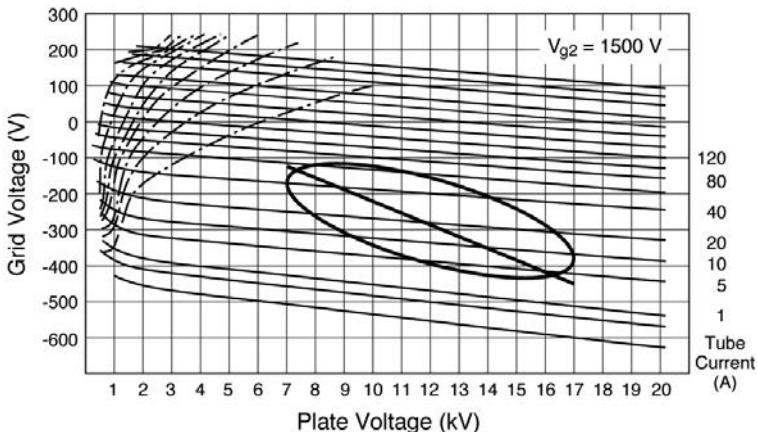


Figure 1: Example of constant-current characteristics for a high power tetrode. If the load is real then the load line is straight, if imaginary the load line is an ellipse. Several iso- $I_a$  lines are shown with current in A indicated to the right.

$I_a$  = anode current;  $V_g$  = control grid voltage;  $V_s$  = screen grid voltage;  $V_a$  = anode voltage;  $\mu_s$  = screen grid amplification factor;  $\mu_a$  = anode amplification factor;  $k$  = geometrical factor of the tube called perveance (Sec.7.3.2.1).

The relative effectiveness of the control grid voltage to anode voltage on the anode (plate) current is called the amplification factor  $\mu$ ,

$$\mu = (\partial V_a / \partial V_g)_{I_a} \quad (1)$$

The screen grid lying between control grid and anode screens the control grid from changes in anode potential and is held at dc potential and zero rf voltage. This makes the cathode-anode current essentially independent of anode voltage, rendering the tetrode a constant current device. Following Child's Law [4] (Sec.7.1.1.1)

$$I_a = k \left( V_g + \frac{V_s}{\mu_s} + \frac{V_a}{\mu_a} \right)^{3/2} \quad (2)$$

For practical design, the behavior of the tube is represented by its "characteristics" provided by the manufacturer. These are families of curves showing  $I_a/V$  relationships at the various electrodes. Fig.1 displays an example [5]. The effects of interelectrode capacitances must be included in the circuit models at high frequencies.

In accelerator applications the residual dependence of current on anode voltage is important as it contributes to the beam impedance of the cavity/amplifier set.  $I_a$  will be  $\propto$  beam current and

$V_a \propto$  the voltage seen by the beam. The dynamic anode resistance of the tube is defined as

$$r_a = (\partial V_a / \partial I_a)_{V_g} \quad (3)$$

Although  $r_a$  would be infinite in a true constant current device, in large high voltage tubes with large surface areas and spacings, the finite value of  $r_a$  can significantly lower the beam impedance of the cavity/amplifier set. Reduction by a factor of 2-3 is practical.

The transconductance is

$$g_m = (\partial I_a / \partial V_g)_{V_a} \quad (4)$$

and its relation to the other parameters is

$$g_m = \mu / r_a \quad (5)$$

Since one usually wants low  $r_a$  and high  $\mu$ ,  $g_m$  can be a useful measure of the performance of the tube. The dissipated power and  $g_m$  rise rapidly with average current in the tube. Dynamically controlling the average current by changing the control grid bias can be used to enhance gain and decrease beam impedance at isolated key points in the accelerator's operating cycle.

The presence of the screen grid greatly enhances the stability of the tetrode. By providing an electrostatic shield between the anode and control grid the inter-electrode capacitance is reduced by a large factor ( $\sim 20$  for high-power tubes). This capacitance has two main deleterious effects; it

causes spurious feedback around the stage and reduces stability, and as seen from the input (control grid) its value is multiplied by the stage gain (Miller effect) thereby increasing the required drive power. The reduced anode-to-grid capacitance of the tetrode implies that it can be operated in the common-cathode (grid-driven) configuration up to high frequencies ( $\sim 100$  MHz). The common-cathode configuration features high power gain and good impedance isolation between load and driver.

**Amplifier** Control grid bias determines the “class” of operation, A, B, or C. In A, the tube conducts throughout the rf cycle making for good linearity but poor (<30%) efficiency. In C, the tube conducts for less than half the rf cycle making for good efficiency (up to 100%) but high distortion. For simple driving of a high  $Q$  cavity distortion is not important but if rapid modulation or wide dynamic range is required, class C cannot be used. In addition, class C operation does little to decrease beam impedance since the tube impedance is  $\sim \infty$  when cut-off.

Class B operation is a compromise where current flows for half the rf cycle and efficiencies are typically 66%. In this case any amplitude of grid drive will generate some response in the output current. The output wave form is distorted but operation down to a low level is possible. Often in accelerators where good control of rf cavity voltage, over wide dynamic range is necessary, the mode chosen is set between A and B and designated class AB. Sometimes subscripts are included, (AB<sub>1,2</sub>) to indicate whether grid current is (2) or is not (1) flowing.

Detailed analysis of the amplifier’s performance will rely on the characteristics curves, e.g. Fig.1. These plots are provided at various screen voltages and interpolation using the 3/2 power law can be used for specific screen voltages. In order to determine the power efficiency and the input voltage needed for a given load impedance and voltage, one uses a graphical procedure [3] to obtain instantaneous values of anode currents and voltages for a complete rf cycle. First choose a quiescent point at which the tube will operate with zero ac voltage at the grid. This choice will determine the class of operation. Zero quiescent current implies class C or B operation. For a real load the point of maximum tube current and minimum tube voltage is determined. This point and the quiescent point determine a line, called the

*load line*. It extends symmetrically about the quiescent point and is the locus of points over which the anode and grid voltages range in an rf cycle. If the load impedance is not real then the locus will be an ellipse. This follows from the fact that, to a first approximation, the anode current is proportional to the grid voltage. The load ellipse can be written parametrically as

$$E_{\text{anode}}(\omega t) = E_a^q - E_a^0 \sin \omega t \quad (6)$$

$$E_{\text{grid}}(\omega t) = E_g^q + E_g^0 \sin(\omega t + \phi) \quad (7)$$

where  $E_g^q$  is the quiescent grid bias,  $E_g^0$  is the amplitude of rf grid voltage,  $E_a^q$  is the quiescent anode voltage,  $E_a^0$  is the amplitude of the anode’s rf voltage,  $\phi$  is the phase angle of the load impedance (first estimate for the phase of the grid voltage) and  $\omega$  is the radian frequency of the rf drive.

To find the actual load current, and hence the output power, the load ellipse is plotted on the tube characteristic and the anode current  $I_a(t)$  is estimated for each point,  $(E_g, E_a)$ , see Fig.1. Because the tube current is, in fact, a nonlinear function of grid voltage the output current will not be a pure sinusoid. The fundamental and harmonic components of the output current are determined by Fourier decomposition,

$$I_a(t) = a_0 + 2 \sum_{n=1}^{\infty} a_n \cos \Omega + b_n \sin \Omega \quad (8)$$

where  $\Omega = 2\pi nt/T_0$ ,  $T_0$  = rf period, and the Fourier coefficients are calculated from

$$a_0 = \frac{1}{N} \sum_{n=1}^N I_a \left( n \frac{T_0}{N} \right), \quad N = \text{no. of } I_a \text{ samples}$$

$$a_n, b_n = \frac{1}{N} \sum_{n=1}^N I_a \left( n \frac{T_0}{N} \right) \cos \Omega, \sin \Omega$$

The fundamental plus dc output current is then

$$I_a^1 = a_0 + A \sin(\omega t + \theta) \\ \theta = \tan^{-1}(a_1/b_1), \quad A = 2\sqrt{a_1^2 + b_1^2}$$

Since the actual phase of output current with respect voltage is  $\theta$ , the procedure may need to be iterated with a new estimate for the phase of the grid voltage,  $\phi$ . Finally, the fundamental output power is

$$P_{\text{rf}} = (1/2) A E_a^0 \cos(\theta) \quad (9)$$

The total input power is

$$P_{\text{in}} = a_0 E_a^q \quad (10)$$

The anode dissipation,  $(P_{\text{in}} - P_{\text{rf}})$ , must be within the cooling capacity of the tube. The conversion efficiency is

$$\epsilon = P_{\text{rf}} / P_{\text{in}} \quad (11)$$

**Practical considerations** Although the tetrode is inherently quite stable due the anode-grid isolation afforded by the screen grid, at high frequencies it may nevertheless be necessary to “neutralize” the amplifier by supplying some negative feedback between the anode and control grid. This is usually done by coupling a small fraction of the anode voltage back into the grid circuit with appropriate polarity to cancel the effect of the anode-grid capacitance. The screen lead inductance naturally creates some feedback of the correct polarity for neutralization. The frequency dependence of this feedback is strong and at some high frequency ( $>100$  MHz) the tube will self-neutralize. Operation above the self-neutralization frequency requires special precautions.

In high power applications, there may be some electron current flow into the control or screen grids for a portion of the rf cycle. This happens when the control grid potential becomes positive with respect to the cathode and when the anode voltage becomes negative with respect the screen grid. This real current flow will cause power dissipation in the grids which are weakly cooled. Tube characteristics include data on the magnitude of grid currents which must be analyzed in a similar way to that described above for anode dissipation. The maximum dissipation at the grids is typically a factor of 100 or more below that at the anode. One technical but significant detail resulting from the real grid-current flow is the effect it may have on the ancillary power supplies that provide dc bias at the grids. Since this current will create a voltage opposite to the dc bias the supplies must be able to sink the current. Loss of regulation in the bias supplies may lead to a runaway situation where reduced bias causes more grid current and so on.

High power tetrodes must be protected from catastrophic damage due to internal arcing. A fast crow-bar circuit must divert the anode supply current and dump the stored energy in the power supply in less than a few  $\mu\text{s}$ 's.

**Coupling to the cavity** The amplifier must be coupled to the cavity through the appropriate impedance transformation in order to maximize the power flow and efficiency. If a wide frequency

range is needed, such as with ferrite-tuned cavities, then the tetrode should be located close to the cavity.

Also when rf feedback around the amplifier and cavity is used then the amplifier must be close to minimize delay. Direct coupling (via a dc blocking capacitor) to the cavity gap is possible at low frequencies when the cavity impedance is low ( $10 \text{ k}\Omega$ ) and/or multiple gaps are connected in parallel. If the dc bias lead to the anode is routed through a ferrite stack then strong coupling can be achieved without the need for a dc blocking capacitor [8]. Multiple loops through a stack or a loop through a partial stack can give various impedance ratios.

For single-frequency, high  $Q$  cavities an inductive loop is usually used to match the cavity gap impedance to a  $50 \Omega$  transmission line. The size and orientation of the loop are chosen to transform the impedance at the gap to the impedance of the transmission line [9]. For beam-loaded cavities the loop may be over-coupled to the bare cavity in order to effect a good match during the beam pulse. At the amplifier a tank circuit, filter, or cavity will be used to match the tetrode output impedance to the transmission line. A lumped-element filter can be used up to approaching 100 MHz, while for higher frequencies a coaxial cavity containing the anode of the tetrode avoids the spurious resonances of lumped elements. Sometimes the tetrode must be protected from high reflected power which may cause damage from over voltage or over current. This may happen during beam transients such as at injection or during off-resonance operation. Also, cavity breakdown during conditioning causes 100% reflected power. A ferrite circulator can be used above  $\sim 100$  MHz to isolate and protect the amplifier.

## References

- [1] F. E. Terman, Radio Engineers' Handbook, McGraw-Hill (1943)
- [2] L.J. Giacoletto, Electronics Designers' Handbook, second edition, McGraw-Hill (1977)
- [3] Care and Feeding of Power Grid Tubes, Staff of Varian Eimac, San Carlos, California (1982)
- [4] C.B. Child, PR 32 (1911) 492
- [5] Model TH573, Thomson Tubes Electroniques, 38, rue Vauthier, B.P.305, F-92102 Boulogne-Billancourt Cedex, France

- [6] R. Garoby, Lecture Notes in Physics 400, Springer-Verlag (1990)
- [7] J.M. Brennan, PAC95, p.1489
- [8] R.T. Sanders et al, PAC91, p.681
- [9] E.L. Ginzton, Microwave Measurements, McGraw-Hill (1957), p.396

### 7.3.4 Inductive Output Tube

*J.M. Brennan, S. Belomestnykh, BNL*

The Inductive Output Tube (IOT) [1, 2] combines features of the klystron (Sec. 7.3.2.1) and the tetrode (Sec. 7.3.3). It is sometimes called a klystrode [3]. The IOT and klystron are similar in that they both employ a magnetically focused electron beam, the output power is extracted via a resonant cavity that is excited by currents induced by the electron beam, and that the electron beam is intercepted by a collector thus enabling capability of high power handling. They differ in how the electron current is modulated. In the IOT a control grid located very close to the cathode regulates the instantaneous electron current. In the klystron the electron current modulation is created by beam bunching due to velocity modulation of the electrons, imposed by an input and subsequent gain cavities. The effect of this difference is that the IOT's high conversion efficiency at maximum output power (up to 75%) will drop much slower than klystron's efficiency over a wide range of output power. Because the instantaneous electron current is modulated by the DC-biased control grid, the IOT operates in a class B or C regime. Hence, the DC component of the electron beam current scales with the rf output, as in a tetrode. Whereas in a klystron the efficiency falls rapidly at reduced output power as the DC component of the electron current is fixed. To further improve the IOT's efficiency, multi-staged depressed collector devices were developed.

Structurally, the IOT has four major assemblies: gun, anode, output gap and collector. With the exception of the grid, the gun assembly is similar to that of a klystron. The gap between the grid and the cathode is very small and they both have concave shape to facilitate focusing of the electron beam. Since the grid has to maintain its physical shape while operating at about 1000°C, the choice of the grid material is critical. Pyrolytic graphite is used because of its ruggedness even at very high temperatures. The pyrolytic graphite grids have excellent thermal and

mechanical properties. The input circuit of the IOT (input cavity) is external to the device vacuum envelope and is very similar to that of a power grid tube. Because there are no intermediate gain cavities, IOT is a shorter device than the klystron, but has lower gain, typically  $\leq 25$  dB. Therefore the tube requires relatively high power solid state drive amplifier.

The frequency range for which the IOT is applicable spans approximately 0.3 to 1.5 GHz. The major commercial market for the IOT is UHF television transmission where the efficiency is important for economics and the good linearity yields low distortion of complex signals. The reduced size and cost of the IOT also make it appealing in industrial and accelerator applications [4]-[6]. For example, at several storage ring based light sources IOTs were chosen to drive 500-MHz normal- and superconducting rf cavities [7]-[9]. 1300-MHz devices are used predominantly for CW linac cavities [10]-[13]. The linearity of IOT is beneficial for the feedback loops around the amplifier that compensate beam loading. The loop gain is constant for even large modulations of drive power. However, if IOT is operated in a class C regime, it becomes very non-linear at low power levels, which may present problems for feedback loops [14]. The sensitivity to DC power supply fluctuations is much less than that of the klystron because the device is short, and the transit time of the electron beam much less. In this regard the IOT resembles the tetrode. The output coupling cavities are usually (except above 1 GHz) external to the vacuum of the IOT. Often, two output cavities are stager-tuned to achieve 9 MHz bandwidth for TV. Typical power range is 30 kW CW and 100 kW peak. High power units have been developed for special applications [15]-[17]. The tube can be pulsed via the grid bias so a high power modulator is not required for pulsed operation. A crowbar or fast power interrupter must protect the tube from arcs. Manufacturers include: e2v technologies, CPI/Eimac, L-3 Electron Devices, and Thales Electron Devices.

## References

- [1] A.V. Haeff, Electronics, pp. 30-2, Feb. 1939 and Proc. IRE (1940) p. 152
- [2] Care and Feeding of Power Grid Tubes, Prepared by the Staff of CPI/EIMAC, San Carlos, CA, 2003

### Sec.7.3: ACCELERATION

- [3] D.H. Preist, M.B. Shrader, Proc. IEEE vol. 70, no. 11 (1982)
- [4] A. Zolfaghari, P. MacGibbon, B. North, EPAC04, p.1093
- [5] H. Bohlen et al, EPAC04, p.1111; LINAC04
- [6] A. Wheelhouse, E. Sobieradzki, PAC05, p.3883; A. Wheelhouse, EPAC06
- [7] J. Alex et al, EPAC 2004; M. Jensen et al, PAC 95; EPAC 2006
- [8] A. Fabris et al, EPAC02; PAC07, p.2080
- [9] P. Sanchez et al, EPAC 2008; F. Perez, P. Sanchez, A. Salom, IPAC10, p.2615
- [10] J.F. Orrett et al, EPAC06, p.1382; P. McIntosh et al, SRF09
- [11] S. Belomestnykh et al, EPAC06, p.472; LINAC 2008
- [12] H.-G. Hoberg et al, SRF09, p.679
- [13] M. Marks et al, IPAC10, p.4011; A. Wheelhouse et al, IPAC10, p.3999
- [14] A. Nemann, LLRF 2011 Workshop
- [15] M. Shrader, D. Preist, PAC85
- [16] D. Preist, M. Shrader, IEEE Trans. Electron Devices, vol. 38, no. 10 (1991)
- [17] D. Rees et al, PAC95, p.1521

#### 7.3.5 Drift Tube Linacs

*J.M. Potter, JP Accelerator Works*

Beam physics defines the accelerator structure requirements in terms of EM field distributions. The rf and mechanical engineering address the practical realization of the physics design. The basic rf properties of the structure are the resonant frequency, the rf power losses, the coupling to the rf source and the structure tuning. The mechanical properties are vacuum, cooling, alignment and structural integrity. All of these properties must be maintained in spite of aging of components with thermal and vacuum cycling.

Modern ion linear accelerators use a variety of structures to accelerate the beam from the ion source to its final energy. When the LAMPF (now LANSCE) accelerator was built in the early 1970s the DTL bridged the gap between the 750 keV ion source and the 100 MeV and up CCL. The development of the RFQ accelerator in the early 1980s reduced the injection energy requirement to 25 to 70 keV and accelerated the beam to 2 to 2.5 MeV, eliminating the lowest energy part of the DTL which is also the least efficient part and the most difficult to build. In the mid-90's a new accelerator development, the CCDTL [1], was used to reduce the high energy needed from the DTL

and increase the injection energy into the CCL. The structure is a hybrid of CCL and DTL that combines the resonant coupling of the CCL with short cavities containing one or more drift tubes.

The post-stabilized or post-coupled, as it is sometimes called, DTL is a structure that makes it possible to achieve a good approximation to the physics design with readily realizable construction tolerances and good long term stability [2, 3, 4]. Designers who are contemplating building a modern Drift Tube Linac are advised to understand the significant benefits of the post-stabilized DTL. Figure 1 shows the salient features of the post stabilized DTL looking through a transparent outer wall.

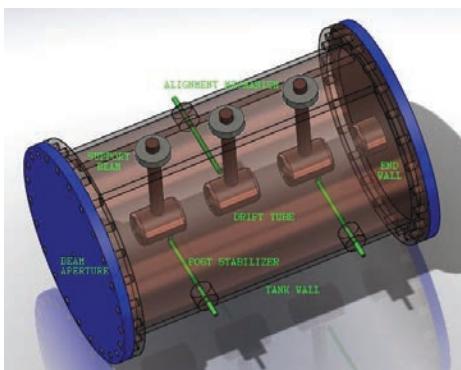


Figure 1: Features of the post stabilized DTL.

The fields in the gaps between drift tubes have the same phase so the cavity is said to operate in the capacitively loaded  $\text{TM}_{010}$  mode.

**Frequency choice** The choice of operating frequency for a DTL structure is governed by two basic principles, the availability of a suitable rf power source for a particular frequency and the required beam aperture. Generally high current applications require larger beam holes and lower operating frequency for good efficiency. DTLs have been proposed for frequencies from 80 to 800 MHz and beam currents up to 100 mA c.w.

Table 1 lists the nominal range of DTL parameters for proton acceleration and the parameters of the SSC and Los Alamos (LAMPF) DTLs as examples. The SSC DTL was designed about 1991 and the Los Alamos DTL was designed in 1967. The difference is made possible by the use of permanent magnet quadrupoles (PMQs) for beam focusing in the SSC DTL rather than the

Table 1: Range of DTL features.

| Quantity                            | Range     | SSC   | SSC Upgrade | LAMPF   |
|-------------------------------------|-----------|-------|-------------|---------|
| Approx. date                        | 1970-2000 | 1990  | 2000        | 1970    |
| $f$ [MHz]                           | 80-800    | 427.6 | 428.3       | 201.25  |
| $\hat{I}_{beam}$ [mA]               | 10-100    | 25    | 25          | 15      |
| $\bar{I}_{beam}$ [ $\mu$ A]         | 2-1000    | 25    | 1000        | 1000    |
| $E$ [MV/m]                          | 1-8       | 4.6   | 4.6         | 2.5     |
| $T_{inj}$ [MeV]                     | 0.75-2.5  | 2.5   | 2.5         | 0.75    |
| $T_{out}$ [MeV]                     | 4-200     | 70    | 33.2        | 100     |
| $g_0 \cdot L$ [T]                   | 1-6       | 4.64  | 4.64        | 1.4-2.2 |
| $a$ [mm]                            | 5-50      | 16    | 16          | 15-30   |
| $D_{tank}$ [cm]                     | 20-250    | 42    | 42          | 88-94   |
| No. Tanks                           | 1-9       | 4     | 2           | 4       |
| $D_{drift tubes}$ [cm]              | 4-40      | 8     | 8           | 16-18   |
| $\Delta L_{gap}$ [ $\beta\lambda$ ] | 1-2       | 1     | 1           | 1       |
| $L_{tot}$ [m]                       | 2-100+    | 24.44 | 11.1        | 61.74   |
| $P_{rf}$ [MW-pk]                    | 0.2-20    | 6.4   | 2.5         | 8.4     |
| $P_{beam}$ [M-pk]                   | 0.04-20   | 1.7   | 0.83        | 1.7     |
| $P_{beam}$ [kW-avg]                 | 0-1000    | 1.7   | 33          | 100     |
| $\Delta T/T$                        | -         | 0.006 | .006        | 0.01    |
| $\varepsilon$ [ $10^{-6}$ m·r]      | -         | 0.63  | 0.63        | 1.5     |

where  $\hat{\cdot}$  designates peak value, bar-over the average value,  $a$  the beam aperture diameter,  $g \cdot L$  the integrated quadrupole gradient,  $E$  the accelerating gradient and  $T$  the kinetic energy.

electromagnetic quadrupoles used in the LAMPF (now LANSCE) DTL. The first two tanks of the SSC linac have been upgraded from 0.1% duty factor to 4% duty factor to make an isotope production facility in Denton, TX. [5] The modified SSC accelerator parameters are included in the comparisons of Table 1.

**Resonant frequency** Principal sources of frequency error in a DTL: machining tolerances of tank I.D., machining tolerances of drift tube length, and inaccuracies in calculating the frequency with acceptable mesh sizes. Laboratory measurements of frequency must be corrected for the frequency shift due to the dielectric constant of air and the difference between laboratory temperature and the operating temperature. Heating from high power rf produces thermal gradients in the structure that lower the operating frequency more than estimated by using the average temperature. The frequency also is lowered as the rf power coupling is increased when the DTL is coupled to a wave guide through a coupling iris.

Achieving the correct frequency requires a hierarchy of tuning mechanisms to correct for these factors. The tank inside diameter is increased to

insure that the worst case frequency is lower than desired. A bulk tuner in the form of tuning bars running the length of the structure decreases the volume of the magnetic field region of the tank and raises the frequency. Slug tuners provide a fine adjustment of the frequency. Often, one or more slug tuners are motor driven to enable the frequency to be phase-locked to the rf source. Alternatively, variable-field ferrite-loaded transmission lines can be as an electronic tuning mechanism.

The frequency corrections and power losses for the drift tube stems and post stabilizers can be calculated by modern 3-D electromagnetic field programs such as HFSS™, Microwave Studio™ or Soprano™. There are many such programs available. The designer should carefully investigate which best fits his needs, budget and available computing power. Figure 2 shows the E-field on the mid-plane of the DTL shown in Figure 2, as calculated by HFSS.

The model was made periodic by including half stems and half post stabilizers on the end walls to facilitate determining the mode spectrum as a function of the post-stabilizer-to-drift-tube

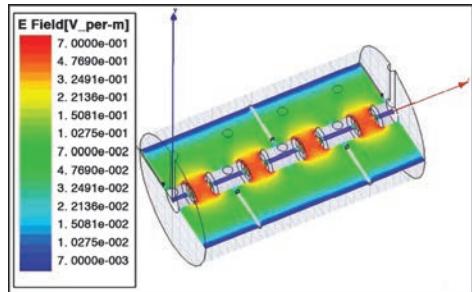


Figure 2: Electric field plot of accelerating mode for DTL test model.

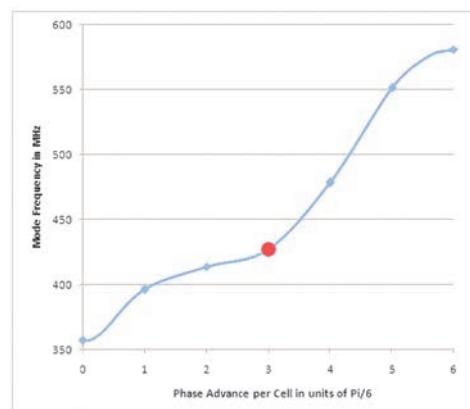


Figure 3: Post stabilized DTL model calculated mode spectrum.

gap. Figure 3 shows the calculated mode spectrum with the stopband closed and a finite group velocity at the accelerating mode (red dot) indicating resonant coupling. The modes were calculated with short circuit boundaries at the ends. The coupler mode was calculated with open circuit end boundaries.

Figure 4 shows the on axis accelerating field,  $E_z(z)$ , for the accelerating mode. The eigenvalue solver took four hours to find seven modes to a precision of 0.05% ( $\pm 200$  kHz). However, the absolute accuracy is dependent on how many-sided a polygon is used to simulate circular cross sections. The unevenness of the plot points out some of the difficulties of using 3-D codes. Even with 90,000 tetrahedrons and energy density mesh optimization the field plot shows the effect of the mesh structure in the differences between what

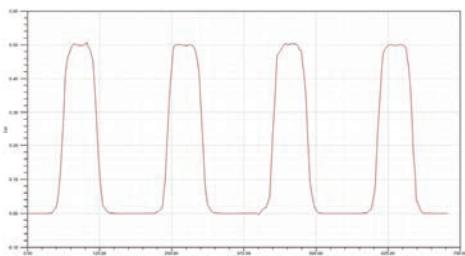


Figure 4: On axis  $E_z$  field for accelerating mode.

should be identical fields in each gap. For periodic structures HFSS has a Master/Slave boundary condition that allows calculation of mode frequencies anywhere in the passband. The calculations were done on a Dell 690 with 16 GB RAM, two quad core 2.66 GHz processors and a 64 bit operating system.

These results emphasize that even with modern 3-D codes and powerful workstations the usefulness of the calculations depend on the skills of the designer. The old standby program, SUPERFISH [6, 7], is still available and can still be used to estimate frequency corrections and power losses. For more information on the use of SUPERFISH in DTL design see previous editions of the Handbook [8].

DTL structures are large and invariably rely on bolted mechanical rf joints so the accelerator can be fabricated from units of manageable size. Poorly designed rf joints can create serious maintenance problems. Damage due to local heating can quickly destroy an rf joint. There is a run-away effect in a bad joint. When an area is damaged by overheating and arcing additional stress is placed on the adjacent contact areas. A bad rf joint can literally "come unzipped". The best joints are welded or brazed, but this is not always practical. If a mechanical rf joint is required care must be taken in the selection of joint materials, especially in regions of high rms current density. Table 2 lists some common joint materials and their range of applicability.

In some instances the rf joint material can also serve as a vacuum seal. Additionally, the C-seal can have enough compliance to allow positional adjustments of 0.25 mm or so, depending on its height. However, it is not advisable to combine position adjustment, rf contact and vacuum sealing in one device. The conflicting requirements for the three functions are such that a C-seal can

Table 2: RF joint materials.

| Material               | Application     |
|------------------------|-----------------|
| Sn-Fe mesh (Techknit™) | low current     |
| Be-Cu band (Multilam™) | medium current  |
| Inconel C, Cu plate    | high current    |
| Crushed wire           | not recommended |

be used for any combination of two functions but not all three. The two requirements for a good rf joint are a high conductivity material and a high pressure contact with a strong restoring force.

**Accelerating field gradient** In early DTL structures the on-axis electric field gradient was the order of 1 MV/m throughout the structure. More recently structures have been designed with much higher fields, up to 8 MV/m. The limiting factor, other than the increased rf power required to achieve the design energy, is breakdown in the drift tube gaps due to high peak surface fields. Peak surface field levels up to 1.8 times the Kilpatrick criterion [9] (Sec. 1.6.18) can be used. This is about 36 MV/m at 428 MHz. The SSC DTL was designed to operate at 4.6 MV/m with a peak surface field of about 26 MV/m, or about 1.3 Kilpatrick. The SSC design is considered to be conservative.

To facilitate capturing the beam from an RFQ by a high gradient DTL, it is common to ramp the on-axis field from a value comparable to the RFQ accelerating field up to the DTL maximum field in a structure is called a ramped gradient DTL or RGDTL. The ramp can be linear over the entire length of a tank. The ramped gradient section can be combined with a flat field in a single tank if the transition is smooth.

**Post Stabilizers** The post stabilizer is a resonant coupling device that performs an important function similar to the coupling cavities in a CCL, reducing the sensitivity of the field distribution to tuning errors by several orders of magnitude and making it easier to build and tune a DTL.

Empirical evidence from DTL accelerators and scale models indicate that the spacing between the drift tube and the wall must lie between  $\sim 0.90$  and  $\sim 1.03 \lambda/4$  [10]. The requirement for high efficiency suggests making the drift tube diameter as small as possible, limited by the PMQ diameter. The recommended drift-tube-to-wall distance is  $0.95 \lambda/4$ , which effectively determines the tank inside diameter once the drift tube diameter is chosen. Fine adjustment of the

DTL cells is then determined by the drift tube length.

3-D codes calculations including the post stabilizers should be used to verify the choice of tank ID and drift tube OD.

A properly tuned stabilizer forces the voltage on the drift tube to be zero at the stabilizer's electrical center. If the stabilizer is located at the midpoint of a drift tube with an eccentric tab rotated straight up or down the voltage at the two ends of the drift tube are equal and opposite. The rotation of the tabs is used to force the electric field distribution to follow the design with an acceptable tolerance. The optimum design procedure for a RGDTL is to detune the end gaps in opposite directions. *i.e.* one gap is made smaller the other larger, to force the field to be naturally tilted. The post stabilizers can then be centered on the null voltage location for each drift tube. Thus, the stabilizers are not forcing the tilt which results in large stabilizer currents.

**Tuning** A detailed tuning procedure for the post stabilized DTL can be found in previous editions of the Handbook [8] and in reference [13].

**Tank to tank spacing** In multi-tank DTLs the tank spacing is generally 1 or 2 times  $\beta\lambda$ , and the periodicity of the PMQ focusing system is continued through the inter-tank region. This space is often used for vacuum valves to isolate the tanks for maintenance, for diagnostics such as beam current transformers and position sensors and possibly for steering magnets. If the distance is longer, there will be a problem getting the beam into the acceptance of the following tank due to phase spreading of the beam from space charge forces. A spacing of  $3 \beta\lambda$  is possible if the gap positions at the ends of the adjacent tanks are manipulated to decrease the beam phase spread at the exit and increase the phase acceptance at the entrance. This technique increases the complexity of the DTL and makes it harder to design because the gaps cannot be moved arbitrarily without taking steps to insure the uniformity of the accelerating gradient [11].

**PMQ alignment** With a post-stabilized DTL, the most critical mechanical adjustment is the transverse alignment of the drift tubes. Error analyses have shown that the PMQ axes must be aligned to a smooth curve (not necessarily a straight line) within about  $50 \mu\text{m}$  [11]. A useful technique for aligning the PMQs is the pulsed taut wire [12]. If a thin wire is strung through the

### Sec.7.3: ACCELERATION

beam aperture and pulsed with a current, a transverse acoustic wave is induced the wire. The wave amplitude is proportional the displacement of the PMQ from the axis is separated in time proportionally to the magnet spacing.

Calculating the fields of a complete DTL tank is still taxing the capabilities of the PC in terms of memory required and time to complete calculations. There are techniques for partitioning the calculations so that it is practical to use a PC for DTL design. For a detailed discussion see "Field Stabilization with Post Couplers for DTL Tank1 of Linac4" [13].

### References

- [1] J.H. Billen et al, 1994 Linac Conf. p. 341
- [2] D.A. Swenson et al, Proc. 6th Int. Conf. High Energy Acc., CEAL-2000 (1967), p. 167
- [3] L.W. Alvarez, PR 70 (1946) 799
- [4] L.W. Alvarez et al, RSI 26 (1955) 210
- [5] International Isotopes, Inc. succeeded by Trace Life Sciences, Denton, TX.
- [6] K. Halbach, R. Holsinger, PA 7 (4) (1976) 213
- [7] A PC version of SUPERFISH is available from the LANL Code Group, <http://laacg1.lanl.gov/laacg>
- [8] J.Potter in Handbook of Accelerator Physics & Engineering, 1st ed, 3rd printing, World Scientific (2006)
- [9] W.D. Kilpatrick, U. Calif. Rad. Lab. Report UCRL-2321 (1953)
- [10] J. Billen, LANL memorandum AT-1:84-74 (1984)
- [11] K.R. Crandall, private communication
- [12] C. Fortgang et al, 1990 Linac Conf., p. 426
- [13] N. Alharbi et al, CARE-Note-2006-012-HIPPI

#### 7.3.6 Normal Conducting $v = c$ Linac Structures

*G.A. Loew, SLAC*

Symbols:  $\lambda_0$  = free space wavelength;  $v_p$  = phase velocity;  $v_g$  = group velocity;  $\beta$  = propagation constant or cavity coupling factor;  $\beta_n$  = propagation constant of  $n$ -th space harmonic;  $r$  = shunt impedance per unit length;  $s$  = elastance per unit length;  $\alpha$  = attenuation per unit length;  $\ell$  = structure length;  $\bar{E}$  = average accelerating field or gradient;  $E_s$  = peak surface electric field;  $P$  = peak rf power;  $P_0$  = peak input rf power;  $t_F$  = structure filling time, TW only;  $T_0$  = structure filling time, SW only;  $w_{ST}$  = stored energy per unit length;  $w_{\ell,ST}$  = energy stored per

structure;  $k_0$  = longitudinal, fundamental mode loss factor;  $\sigma_z$  = bunch length;  $I$  = peak beam current;  $V$  = integrated voltage seen by particle of unit charge;  $\eta$  = efficiency;  $\eta_{ST}$  = ratio of energy stored in structure to energy supplied by source;  $d$  = disk-loaded waveguide periodic length;  $2a$  = disk-loaded waveguide iris diameter;  $2b$  = disk-loaded waveguide cavity diameter;  $t$  = disk thickness.

The SLAC linac parameters are typical of S-band constant gradient structures:

|   |                       |
|---|-----------------------|
| frequency [MHz]   | 2856                  |
| mode  | $2\pi/3$              |
| length  | 84 cells + 2 couplers |
| $t$ , disk thickness [inch]                             | 0.23                  |
| iris lip radius [inch]                                  | 0.122                 |
| $2a$ , iris diameter [inch]                             | 0.7517 - 1.032        |
| $2b_{max}$ , max cavity inner diameter [inch]           | 3.28                  |
| $t_f$ , filling time [ms]                               | 0.83                  |
| $\tau$ , attenuation [neper]                            | 0.57                  |
| $v_g/c$ , group velocity/ $c$                           | 0.0204 - 0.0065       |
| $r$ , shunt imped. [ $M\Omega/m$ ]                      | 53 - 60               |
| $r_{avg}$ average shunt impedance [inch]                | 56.5                  |
| $Q$ nominal   | 13,000                |
| $\Delta f/\Delta T$ , temp. coeff. of frequency [kHz/K] | 50                    |

Linac structures can be of the traveling-wave (TW) or standing-wave (SW) type. TW structures can be sub-divided into constant-impedance and constant-gradient structures. There is a variation of the constant-gradient structure in which the taper is adjusted to give a quasi-constant gradient together with a Gaussian distribution of the HOM frequencies for the purpose of detuning wake-fields to ameliorate coupled bunch effects. Most of the parameters used in the design of these machines have common definitions.

The formulae presented in this section have been derived and published in a number of books and reports over the last 40 years. There is no single place where all of them can be found. [1, 2, 3] are useful for overview. Discussions on specific topics: beam loading [4, 5, 6, 7, 8], space harmonics [9], measurement techniques [1, 3, 10] standing-wave structures [11, 12], structure design for linear colliders [13, 14, 15], linac technology for linear colliders [16], various compendia on linear accelerators [17, 18] and rf separators [19] (Sec.7.2.10).

**Basic relations and scaling with frequency**  
See Tab.1.

Table 1 Frequency scaling

$v_g$  independent of  $\omega$  for direct scaling of modular dimensions. \*\*  $\tau, \bar{E}, v_g$  constant.

| Quantity   | scales as $\omega^{\frac{1}{2}}$ |
|--|----------------------------------|
| $r = E^2 / -(dP/dz)$<br>( $E$ of synchronous space harmonic)         | 1/2                              |
| $Q = \omega w_{ST} / (-dP/dz)$                                       | -1/2                             |
| $r/Q = E^2 / \omega w_{ST}$  | 1                                |
| $s = E^2 / w_{ST}$   | 2                                |
| $k_o = (\omega/4)(r/Q)$  | 2                                |
| $\alpha = -(dP/dz)/2P = \omega/2Qv_g$                                | 3/2                              |
| $\tau = \alpha\ell$  | 0                                |
| $t_F = \int_0^\ell v_g^{-1} dz = (2Q/\omega)\tau$<br>( $r$ constant) | -3/2                             |
| $\ell = v_g t_F$   | -3/2 *                           |
| $P_o \propto \bar{E}^2 \ell/r$                                       | -2**                             |
| $P_o/\ell \propto \bar{E}^2/r$                                       | -1/2**                           |
| $w_{ST}(z) = P(z)/v_g(z)$  | -2**                             |
| $w_{\ell,ST} \propto V^2/\omega(r/Q)\ell$                            | -7/2**                           |
| $T_o = 2Q_L/\omega = 2Q_o/\omega(1+\beta)$                           | -3/2                             |

Space harmonic properties are described by

$$E_z(z, t) = \sum_{n=-\infty}^{+\infty} a_n C J_0(k_{rn} r) e^{i(\omega t - \beta_n z)} \quad (1)$$

where  $a_n$  is the amplitude of the  $n$ th space harmonic,  $C$  is a fixed amplitude factor,  $\beta_n = \beta_0 + 2\pi n/d$ ,  $k_{rn}^2 = k^2 - \beta_n^2$  and  $v_{pn} = \omega / (\beta_0 + 2\pi n/d)$ . For  $n < 0$  the phase velocity of the space harmonic is negative. Note that for  $n = 0$ ,  $k_{r0} = 0$  and  $J_0 = 1$ , which means that for this mode the accelerating field is independent of radius, i.e. it is aberration free.

Exact relationships between the guide properties above and  $a, b, t$  and  $d$  can be obtained from experimental measurements using dielectric rods and beads (Sec.6.16) or by calculation using computer codes such as

| Code      | f-2D        | F-3D        | t-domain |
|-----------|-------------|-------------|----------|
| SUPERFISH | yes         | no          | no       |
| MAFIA     | yes         | yes         | yes      |
| SLAC      | $\Omega$ -2 | $\Omega$ -3 | n.a.     |

Other codes in use are YAP, TRANSVERS and TWAP. The latter was used by P. Wilson and

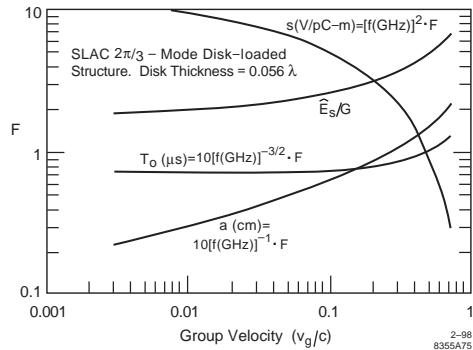


Figure 1: Design Parameters of SLAC type disk loaded waveguide as a function of normalized group velocity.

D. Farkas to prepare the curves of Fig.1 for the design of disk-loaded waveguides of the SLAC S-band structure,  $t = 0.056\lambda$ .

**TW Constant Impedance Linacs**  $\alpha = \text{const.}$   
In steady state (single section)

Define:  $e_1 = 1 - e^{-\tau}$ ;  $e_2 = 1 - e^{-2\tau}$ ;  $e_\alpha = 1 - e^{-\alpha z}$ ,

$$V = \frac{e_1}{\tau} \sqrt{2\tau P_0 \ell r} \cos \theta_0 - ir\ell \left( 1 - \frac{e_1}{\tau} \right) \quad (2)$$

where  $\theta_0$  = particle displacement from wave crest.

$$E = \left( \frac{2P_0 r \tau}{\ell} \right)^{1/2} e^{-\alpha z} - ir (1 - e^{-\alpha z}) \quad (3)$$

$$P(z) = P_0 e^{-2\alpha z} - 2i \sqrt{\frac{P_0 r}{2\alpha}} e_\alpha e^{-\alpha z} + \frac{i^2 r \ell}{2\tau} e_\alpha^2 \quad (4)$$

for bunches on crest.

The power induced by the beam at the structure end is

$$(1 - e^{-\tau})^2 (i^2 r \ell / 2\tau) \quad (5)$$

The energy supplied to the structure is

$$P_0 t_F = \left( \frac{\tau}{1 - e^{-\tau}} \right)^2 \frac{V^2}{\omega \left( \frac{r}{Q} \right) \ell} \quad (6)$$

Energy stored in the structure in the absence of beam is

$$w_{\ell,ST} = \frac{\tau}{2} \frac{e_2}{e_1^2} \frac{V^2}{\omega \left( \frac{r}{Q} \right) \ell} \quad (7)$$

$$\eta_{ST} = \frac{1 - e^{-2\tau}}{2\tau} \quad (8)$$

Normalized peak power required per length

$$\frac{P/\ell}{E^2/r} = \frac{\tau}{2(1-e^{-\tau})^2} \quad (9)$$

In Transient regime (valid for  $0 < t_b < t_F$  where  $t_b$  is the time interval after beam turn on),

$$V = \frac{e_1}{\tau} \sqrt{2\tau P_0 r \ell} \cos \theta_0 - i r \ell F_1 \quad (10)$$

where

$$F_1 = 1 - \frac{1 - e^{-\tau \frac{t_b}{t_F}}}{\tau} - e^{-\tau \frac{t_b}{t_F}} \left( 1 - \frac{t_b}{t_F} \right) \quad (11)$$

If  $t_b \ll t_F$  and the first term is called  $V_0$ , then

$$V = V_0 - \frac{i t_b \ell \omega r}{2 Q} = V_0 - 2 k_0 \ell q \quad (12)$$

This simplified formula can be used for a single bunch or a short train of bunches of total charge  $q$ . It means that the average energy lost by an electron in a bunch is  $k_0 q$ , but the first one loses no energy whereas the last one loses  $2k_0 q$ . This calculation ignores the effect of higher order longitudinal modes which decohere after a few periods of the fundamental mode.

**TW Constant Gradient Linacs**  $\alpha = \alpha(z)$ . Since  $E^2(z) = -r(dP/dz)$  then for constant  $E$ ,  $dP/dz$  must also be constant. Then

$$P(z) = P_0 \left[ 1 - \frac{z}{\ell} e_2 \right] \quad (13)$$

$$\tau = \int_0^\ell \alpha(z) dz \quad (14)$$

and

$$v_g(z) = \frac{\omega \ell}{Q} \frac{\left[ 1 - \frac{z}{\ell} e_2 \right]}{e_2} \quad (15)$$

$$\alpha(z) = \frac{e_2}{2\ell \left[ 1 - \frac{z}{\ell} e_2 \right]} \quad (16)$$

In steady state

$$V = \sqrt{e_2 (P_0 r \ell)} \cos \theta_0 - \frac{i r \ell}{2} \left( 1 - \frac{2\tau e^{-2\tau}}{e_2} \right) \quad (17)$$

$$E = \sqrt{\frac{r P_0}{\ell} e_2} + \frac{i r}{2} \ln \left[ 1 - \frac{z}{\ell} e_2 \right] \quad (18)$$

$$P(z) = P_0 \left[ 1 - \frac{z}{\ell} e_2 \right] G^2(i) \quad (19)$$

where

$$G(i) = \left\{ 1 + \frac{1}{2} \sqrt{\frac{ri^2\ell}{P_0 e_2}} \ln \left[ 1 - \frac{z}{\ell} e_2 \right] \right\} \quad (20)$$

$$P(\ell) = P_0 e^{-2\tau} - 2i \sqrt{P_0 r \ell \frac{\tau^2}{e_+}} e^{-2\tau} + i^2 r \ell \left( \frac{\tau^2}{e_+} \right) \quad (21)$$

where  $e_+ = e^{2\tau} - 1$ . Power induced by beam at end of structure is

$$i^2 r \ell \tau^2 / e_+ \quad (22)$$

and the energy supplied

$$P_0 t_F = \frac{2\tau}{e_2} \frac{V^2}{\omega \frac{r}{Q} \ell} \quad (23)$$

where  $\frac{V^2}{\omega \frac{r}{Q} \ell}$  is the energy stored in the structure in the absence of beam,

$$\eta_{ST} = \frac{1 - e^{-2\tau}}{2\tau} \quad (24)$$

$$\frac{P/\ell}{E^2/r} = \frac{1}{1 - e^{-2\tau}} \quad (25)$$

Transient regime

$$V = \sqrt{e_2 P_0 r \ell} \cos \theta_0 - \frac{i r \ell}{2} T_1 \left( \frac{t_b}{t_F} \right) \quad (26)$$

where

$$T_1 \left( \frac{t_b}{t_F} \right) = \left( 1 - 2 \tau \frac{t_b}{t_F} e^{-2\tau} - e^{-2\tau \frac{t_b}{t_F}} \right) / e_2 \quad (27)$$

For  $t_b \ll t_F$ , the voltage relation is again Eq.(12).

**SW Linacs** SW structures are generally built with  $v_g = 0$  and  $\pi$  phase advance per cavity or, preferably, with high  $v_g$  (5 to 15% c) and  $\pi/2$  phase advance as this gives much less tuning sensitivity to dimensional changes. In the latter case, every other cell has little beam interaction by being very short or off axis. The field build up takes place through rapid multiple reflections back and forth to the coupler with a time constant

$$T_0 = 2Q_0/\omega (1 + \beta) \quad (28)$$

The voltage is [12]

$$V = \frac{2(1 - e^{-t/T_0})}{1 + \beta} \sqrt{\beta P_0 r \ell} - \frac{i r \ell}{1 + \beta} T_2(t, t_b) \quad (29)$$

where  $T_2(t, t_b) = 1 - e^{-(t-t_b)/T_0}$ ,  $t = 0$  when the rf is turned on and  $t = t_b$  when the beam current,  $i$  is turned on. The coupling coefficient,  $\beta$ , is generally set so that the structure is overcoupled ( $\beta > 1$ ) in the absence of beam and matched when the design beam current is on. The responses of the SW structure to rf power and beam turn on are both exponentials with the same time constant. If the actual current is lower than the design current,

one can decrease the rf power at  $t = t_b$  so that the voltage seen by the beam is constant. If  $t_b$  is chosen such that

$$\frac{2\sqrt{\beta P_0 r \ell}}{1 + \beta} e^{-t_b/T_0} = \frac{i r \ell}{1 + \beta} \quad (30)$$

then for  $t > t_b$ ,

$$V = \frac{2\sqrt{\beta P_0 r \ell}}{1 + \beta} \left( 1 - e^{-t_b/T_0} \right) \quad (31)$$

If  $t_b$  is chosen for a perfect match with beam on, then

$$t_b = \tau \ln \frac{2\beta}{\beta - 1} \quad (32)$$

**Energy spread** Electron linacs are often required to meet certain specifications regarding transverse beam emittance (see Sec. 4.3.1) and energy spread. Energy spread can take place (i) within a bunch, and (ii) from bunch-to-bunch. See also Sec. 7.3.1.2.

(i) Within a bunch, energy spread occurs because all electrons do not see the same accelerating field. At low charge (negligible wakefields) a bunch of half rms phase length  $z$  riding exactly on the crest has an energy spread

$$\Delta V/V = 1 - \cos \sigma_z \approx \sigma_z^2/2 \quad (33)$$

At high charge the accelerating field is modified by the wakefield of the bunch. If the bunch is located ahead of the wave crest (in  $z$ ) it is possible, to first order, to compensate the rising slope with the falling wake of a Gaussian bunch. By further bunch shaping, the compensation can be made complete [14].

(ii) Bunch-to-bunch energy spread can be due to several causes such as amplitude, phase and frequency variations from the rf source, injection conditions, temperature drifts, differences in charge among bunches and all types of beam loading. For trains of bunches it takes some time for the longitudinal wakes to decohere and, at high charge, they can create bunch-to-bunch energy differences unless precautions to damp or detune these modes are taken. In the TW case, the fundamental mode beam loading reaches steady state one filling time after injection. The transient can be partially compensated by early injection or by causing the initial rf power to have a time profile so that the accelerator structure is pre-loaded with the field profile caused by steady state beam loading.

## References

- [1] A. Septier, P.M. Lapostolle, ed's, Linear Accelerators, North-Holland (1970)
- [2] G.A. Loew, R. Talman, AIP Proc. 105 (1983)
- [3] R.B. Neal et al, The Stanford Two-Mile Linear Accelerator, Benjamin (1968)
- [4] R.B. Neal, M. L. Report 379, Stanford U. (1957)
- [5] R.B. Neal, M. L. Report 388, Stanford U. (1957)
- [6] R.B. Neal, M. L. Report 513, Stanford U. (1958)
- [7] G.A. Loew, M. L. Report 740, Stanford U. (1960)
- [8] R.B. Neal, M. L. Report 259, Stanford U. (1961)
- [9] L. Brillouin, Wave Propagation in Periodic Structures, Dover (1953)
- [10] W.J. Gallagher, M. L. Report 767, M Report M-205, Stanford U. (1960)
- [11] P.B. Wilson, AIP Proc. 87 (1981)
- [12] R.H. Miller, LINAC 86, p.200
- [13] K.A. Thompson et al, PA 47 (1994) p.65
- [14] G.A. Loew, SLAC-PUB-3892 (1986)
- [15] R.H. Miller et al, Proc. Linac Conf. (1996)
- [16] G.A. Loew, Int. Linear Collider Technical Review Committee Report, SLAC-R-95-471, Stanford U. (1995)
- [17] G.A. Loew, LINAC 76, p.217
- [18] J. Clendenin, L. Rinolfi et al, Compendium of Scientific Linacs, CERN (1996)
- [19] O.A. Altenmueller et al, SLAC Report 17 (1963)

### 7.3.7 Inductively Loaded Accelerating Cavities

*M. Yoshii, KEK*

The rf accelerating systems of low energy proton or ion synchrotrons have used the coaxial type cavities inductively loaded with the magnetic materials, such as Ni-Zn ferrites, Mn-Zn ferrites, amorphous and nanocrystalline soft magnetic materials. Ni-Zn ferrites are the most common materials in tuning cavities with parallel-biasing techniques, which cover frequency below 10 MHz. Mn-Zn ferrites and amorphous metals are used in proton therapy synchrotrons, which are not required for frequency tuning. Recently, the development of magnetic alloys loaded high field gradient rf cavities has made great progress. Nanocrystalline soft magnetic material (Hitachi-Metal FINEMET®) allows the applications for the high intensity proton synchrotrons and/or FFAG synchrotrons.

**Ferrite materials** Ferrite material is being represented by the general formula of  $MFe_2O_4$ ,

where “M” is a divalent metal oxide, for example, manganese oxide, nickel oxide or zinc oxide. MnZn ferrites have permeability of thousands and low hysteresis loss below 1 MHz. Because of the low specific resistance of a few tens or hundreds of  $\Omega\text{cm}$ , the eddy current loss becomes large at high rf frequency. Mostly MnZn ferrites are used in therapy synchrotrons as un-tuned cavities. A tuned-cavity is designed to get high impedance for high accelerating voltages in proton synchrotrons. NiZn ferrites are mainly used with the parallel biasing technique [1]. Relative permeability is typically hundreds and is controlled by external quasi-dc magnetic field to change the inductance for tuning. The frequency bandwidth of ferrite is below 10 MHz, typically followed by “Snoek’s limit” [2]. NiZn ferrites have high electrical resistance and low dielectric constant ( $\epsilon < \sim 10$ ), and Curie’s temperature is around 100°C-200°C. The Q-factor is 10~50. Toroidal cores of 50 cm size are available at maximum and the thickness is  $\sim 25$  mm. Larger cores are difficult to manufacture.

**Magnetic alloy materials (MA)** The magnetic alloy (MA) materials have been also used in ion accelerators or therapy synchrotrons as un-tuned cavities. Wide frequency sweep is required, though an RF voltage doesn’t have to be high. Also, the application development of the MA materials with a high intensity proton synchrotron started with the J-PARC project in 1995 [3, 4]. Especially, nanocrystalline iron alloy, FINEMET® (FT-3M), has high saturation magnetic flux density (1.3T) and high Curie’s temperature (570°C), which results in more stable magnetic characteristics under high  $B_{rf}$  and at high temperature than the ferrite materials. This MA material is based on 18  $\mu\text{m}$ -thick amorphous ribbon and 25~35 mm-wide, which is heat-treated to form nanocrystalline softmagnetic structure. The layer of ribbon must be electrically insulated with a dielectric silica layer, of which thickness is roughly 2  $\mu\text{m}$ . The ribbon tape can be wound into the core over 1m diameter size.

**Structure** Ferrite/MA loaded cavity has the coaxial structure (Figure 1). In case of MA cavity, DC bias windings for tuning are not necessary. The ferrite/MA material is toroidal shape and located outside of beam pipe. Each gap structure can be considered as a shorted quarter-wave length transmission line (coaxial reentrant cavity structure). i.e., the beam pipe becomes an inner

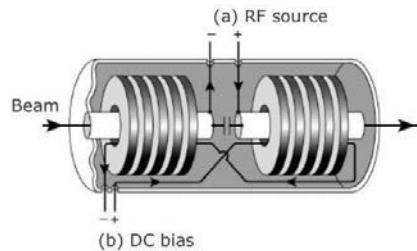


Figure 1: Typical structure of an inductively loaded accelerating cavity.

conductor and the cavity shroud an outer conductor of transmission line. The cavity system has 2 or 3 accelerating gaps, typically. One rf amplifier using triode or tetrode is connected to each gap in parallel and gives the rf power for both generating an accelerating voltage and supplying beam power. Because of relatively low operating frequency, the transit time factor is negligible small. The beam sees the gap impedance in series, but the rf amplifier sees it in parallel. One needs more careful considerations to that point in case of high intensity synchrotrons.

In the ferrite-loaded cavity, the DC biasing (figure 1: (b)) is applied to create an azimuthally dc-magnetic field inside the cores, which changes the ferrite permeability ( $\mu_r$ ) over widely, 100:1. DC bias tuning current is around 2000A-turns at maximum, which is controlled by a servo system (so-called an auto tuning feedback). High bandwidth of the tuning loop is necessary for stability [5]. The figure-8 bias windings are applied for canceling the rf voltages induced at the terminal.

**Cavity Impedances** The rf power dissipation is dominated by the magnetic material cores. The cavity impedance, therefore, is determined by calculating from the geometrical size of cores. The relative permeability of cores;  $\mu_r$ , is shown  $\mu_r = \mu_s' - j\mu_s''$  in the series complex quantity ( $j \equiv \sqrt{-1}$ ). Here, the loss  $\tan\delta = 1/Q$  can be defined as  $\mu_s''/\mu_s'$ . A complex inductance of toroidal core ( $L$ ) is given,

$$\begin{aligned} L &= \mu_o / (\mu_s' - j\mu_s') t \cdot \ln(b/a) / (2\pi) \\ &= L_s' - jL_s'' \\ &= L_s(1 - j\tan\delta). \end{aligned}$$

Where  $t$  is the effective thickness of whole cores,  $a, b$  are an inner and an outer diameters of toroidal core, respectively.

Inductive reactance of core is

$$jL = jL_s + R_s, \text{ where } R_s = \omega L_s / Q.$$

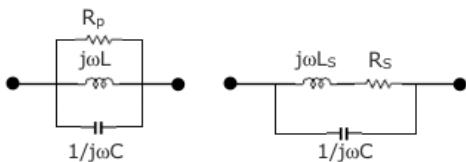


Figure 2: Parallel and Series Expressions of Cavity Equivalent Circuit.

Impedance of magnetic material loaded cavity is equivalently represented as an LCR parallel circuit. The shunt impedance in parallel circuit,

$$R_p = \omega L_p Q = (1 + Q^2) R_s.$$

Moreover,  $R_p$  is proportional to the product of  $\mu'_p Q f$ . Here using the parallel complex permeability ( $\mu'_p$ ),

$$\mu'_p = (1 + 1/Q^2) \mu'_s \text{ or } L = (1 + 1/Q^2) L_s.$$

When  $Q \gg 1$ , it follows that  $\mu'_p \sim \mu'_s$  or  $L \sim L_s$ .

**Properties** The accelerating field gradient performance of the cavity is determined by the ferrite properties and by the cooling technique in case of MA loaded cavity. In generally, the magnetic flux density in an inductive material cannot follow to the change in the rf magnetic field due to magnetic losses.

In this case,

$$B/H = \mu = \mu' - j\mu'' = |\mu| \exp(-j\delta),$$

where  $\mu'$  gives an inductance of material and  $\mu''$  a magnetic loss in the material and  $|\mu|$  is a norm of complex permeability and  $\delta = \arctan(\mu''/\mu')$ . Two materials are compared below,

|         | $\mu'$ | $\mu''$ | Q           |
|---------|--------|---------|-------------|
| NiZn    | 100    | 2       | 50          |
| FINEMET | 2400   | 4000    | 0.6 @ 1 MHz |

According to the Faraday's law, the gap voltage  $V$  is proportional to the time variation of total magnetic flux crossing the cores.

$$V = \omega A B_{rf},$$

where  $\omega$  is frequency, "A" is the cross section of cores and  $B_{rf}$  is a magnetic flux density at an average magnetic length of core.

The specific loss in cores is

$$P_v = \omega B_{rf}^2 / (2\mu_0 \mu'_p Q) [W/m^3].$$

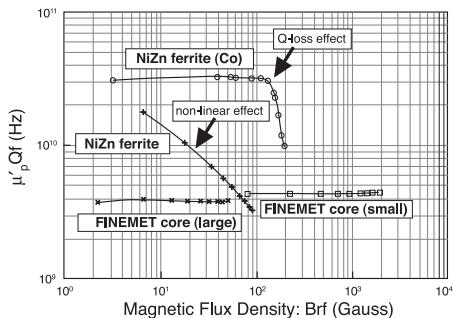


Figure 3: The  $\mu'_p Q f$  properties of N-Zn ferrites and FINEMET cores as a function of  $B_{rf}$ .

The power loss in the toroidal core is concentrated at the inside, because of the  $1/r$  dependence of  $H_{rf}$ . However, the radial dependence of  $B_{rf}$  is reduced when the external biasing for tuning or the cut-core configuration for optimizing Q-value of MA cores are employed.

As shown in Figure 3, most ferrites show non-linear and anomalous Q-loss effects at high  $B_{rf}$  [6]. To avoid thermal runaway of ferrites,  $B_{rf}$  is set to be below 100 gauss and the specific loss in the ferrite is designed to be  $< 0.5 \text{ W/cm}^3$ .

Table 1: Typical RF parameters of various rapid cycling proton synchrotrons; J-PARC RCS is MA loaded cavity, both  $h = 2$  and  $h = 4$  rf voltages are superimposed into the same cavity.

|                      | AGS<br>Booster III | ISIS           | J-PARC<br>RCS     |
|----------------------|--------------------|----------------|-------------------|
| frequency (MHz)      | 1.5-3.5            | 1.4-3.1        | 0.94-1.7          |
| No. of Gaps          | 2                  | 2              | 3                 |
| Accelerating voltage | 45 kV              | 28 kV          | 45 kV             |
| R//                  | 7 kΩ               | 3 kΩ           | 2.5 kΩ            |
| Type of cavity       | ferrite tuning     | ferrite tuning | MA loaded/untuned |
| No. of cores         | 56                 | 70             | 18                |
| Brf (gauss)          | 120                | 103            | 400               |
| Power density (W/cc) | 0.325              | 0.3            | 0.65              |
| Type of core         | Philips 4M2        | Philips 4M2    | FINEMET FT-3M     |
| Size of core         | 500 × 250 × 27.2   | 500 × 300 × 25 | 850 × 375 × 35    |

**Optimum quality factor** Under high beam loading, the RF cavity is driven by the vector sum of generator current and beam current. The beam-induced voltage appears at the accelerating gaps. The RF system is detuned to compensate the loading effect. In case of an untuned cavity, the resonant frequency is fixed and the accelerating frequency is swept to the resonance (below transition energy). It becomes important to alleviate the incoherent tune shift due to space charge. The RF system is also used for increasing the bunching factor by manipulating the longitudinal bunch shape. A 2nd harmonic cavity is employed for this purpose. In J-PARC RCS, the frequency bandwidth of the cavity is designed to cover the two fundamental and 2nd rf frequencies. The optimum quality factor can be chosen over a wide range, 0.6~>30 by using the cut-core and/or the parallel inductance configurations [9]. The multi-harmonic feedforward can effectively compensate the beam loading [10].

**Perspective** Ferrite tuned cavities are the more sophisticated technique. However, designing needs to be conservative, because of the ferrite properties under high magnetic rf field. The MA loaded cavity using the nanocrystalline soft magnetic material is the only system, which can achieve the highest field gradient (25 kV/m). Thus longitudinal space along the ring can be saved. The cavity behaves as a passive load. Reproducible and stable longitudinal operation can be realized by combining with a high-precision digital low level rf control system [10]. Plans for increasing the impedance of the MA cores are underway to obtain the higher field gradient [11].

## References

- [1] I.S.K. Gardner, CERN Acc. School, CERN 92-03, Vol. II, p. 349
- [2] O. Archer, S. Dubourg, PRB 77 (2008) 104440
- [3] Y. Yoshizawa et al, J. Appl. Phys. 64 (1988) 6044-6046
- [4] Y. Mori et al, Proc. of EPAC98, pp. 229-231
- [5] K.W. Robinson, CEA report, CEAL-1010 (1964)
- [6] F. Pedersen, IEEE Trans. Nucl. Sci., NS-32 No. 5 (1985) 2138
- [7] J.E. Griffin, G. Nicholls, IEEE Trans. Nucl. Sci., NS-26 No. 3 (1979) 3965-3967
- [8] Technical Design Report for J-PARC, JAERI-Tech 2003-044
- [9] M. Yoshii et al, PAC07, pp. 1511-1513
- [10] F. Tamura et al, PRST AB12 (2009) 041001 and AB14 051004 (2011)
- [11] C. Ohmori et al, Proc. of IPAC10, pp. 1711-1713

### 7.3.8 Fixed Frequency Cavities

#### 7.3.8.1 Multicell cavities

*W. Schnell, deceased, CERN*

Resonant (standing-wave) cavities are nearly always employed.

Resonant frequency  $f_{mnv} = \omega_{mnv}/2\pi$  of an  $E_{mnv}$  mode ( $\equiv TM_{mnv}$  any mode with  $E_z \neq 0$ ) in a cylindrical cavity (pill box) of radius  $a$  and length  $\ell$ :

$$\omega_{mnv} = \left[ \left( \frac{k_{mn}}{a} \right)^2 + \left( \frac{\pi v}{\ell} \right)^2 \right]^{1/2} \cdot c \quad (1)$$

$k_{mn}$  being the  $n$ -th zero of the  $m$ -th order Bessel function  $J_m(k\rho)$ . For the  $E_{010}$  mode employed for acceleration  $\lambda_0 = c/f_0 = 2.61 a$ .

The choice of frequency is a question of beam dynamics and of minimum cost (transverse size versus number of parts). Frequencies often used are 350, 500 MHz and higher.

Beam-induced higher  $E$ -modes generate energy spread and deflection. Field amplitudes [1]:

$$\begin{aligned} E_z &= E_0 J_m(k\rho) \cos(m\varphi) \cos\left(\frac{\pi v z}{\ell}\right) \\ E_\rho &= -E_0 \frac{\pi v}{k\ell} J'_m(k\rho) \cos(m\varphi) \sin\left(\frac{\pi v z}{\ell}\right) \\ E_\varphi &= E_0 \frac{m\pi v}{k^2 \ell \rho} J_m(k\rho) \sin(m\varphi) \sin\left(\frac{\pi v z}{\ell}\right) \\ H_z &= 0 \\ H_\rho &= -\frac{i\omega_0}{c} \frac{E_0 m}{Z_0 \rho k^2} J_m(k\rho) \sin(m\varphi) \cos\frac{\pi v z}{\ell} \\ H_\varphi &= -\frac{i\omega_0}{c} \frac{E_0}{Z_0 k} J'_m(k\rho) \cos(m\varphi) \cos\left(\frac{\pi v z}{\ell}\right) \end{aligned}$$

with  $Z_0 \equiv \sqrt{\mu_0/\epsilon_0}$ . The indices  $m, n, v$  refer to coordinates  $\varphi, \rho, z$ . Mostly  $\ell = \lambda_0/2$ . Stored energy  $W$  (volume integral of  $\epsilon_0 \frac{E^2}{2}$ ) is computed from  $J_m(ka) = 0$  and

$$\begin{aligned} \int_0^a \rho J_p^2(k\rho) d\rho &= \frac{a^2}{2} \left[ J_p'^2(ka) + \left( 1 - \frac{p^2}{k^2 a^2} \right) J_p^2(ka) \right] \\ J'_m(x) &= J_{m-1}(x) - \frac{m}{x} J_m(x) = -J_{m+1}(x) + \frac{m}{x} J_m(x) \\ 2 \left[ \frac{m^2}{x^2} J_m^2(x) + J_m'^2(x) \right] &= J_{m-1}^2(x) + J_{m+1}^2(x) \end{aligned}$$

Practical cavities require beam apertures and auxiliaries. They often have longitudinal sections

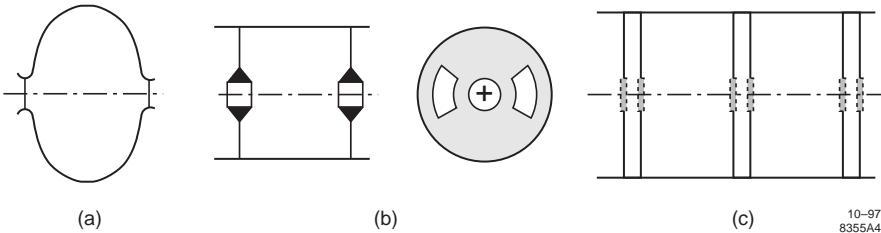


Figure 1: (a) Spheroidal cavity shape, (b) enhanced shunt impedance with cell to cell coupling, (c) a bi-periodic structure.

like Fig. 1(a) or (b). The re-entrant shape (b) offers somewhat larger shunt impedance  $R$  if the beam aperture can be made much smaller than  $a$ . The spheroid shape (a) accommodates a large aperture for moderate loss in  $R$ . More radical departures from the pillbox have been proposed and occasionally employed. Field computations in actual cavities can be done e.g. by MAFIA [2].

Peak accelerating voltage is

$$V_0 = \int_{-\ell/2}^{+\ell/2} E_z(0,0,z) \cos\left(\frac{\omega z}{\beta c} + \psi\right) dz \quad (2)$$

where the particle velocity  $\beta c = c$  in most cases and the maximizing phase angle  $\psi = 0$  for symmetry around  $z = 0$ . Accelerating gradients  $V_0/\ell \approx 2V_0/\lambda_0 \leq 1$  MV/m, limited by cooling problems and cost of power; rf conditioning ( $\sim 100$  h) is always required to overcome multipacting barriers.

Shunt impedance per cavity (the resonant resistance of an equivalent RCL circuit) is defined as

$$R = QZ = \frac{V_0^2}{2P} \quad (3)$$

where  $P$  is the energy loss per unit time. (The classical definition, much used for linacs, is twice that!)

The quality factor is defined as

$$Q = \omega W/P \quad (4)$$

In a free oscillation the stored energy  $W$  decays to a fraction  $1/e$  in  $Q$  radians, the field in  $Q/\pi$  cycles. For given shape the unloaded  $Q$  (due to wall losses) scales with  $f^{-1/2}$ . As a rule of thumb,

$$Q \approx 2 \frac{\text{cavity volume}}{\text{skin volume}} \quad (5)$$

For the  $E_{010}$  pillbox with  $\ell = \lambda_0/2$ ,

$$Q = 0.215 \lambda_0/\delta, \text{ with } \delta = (\pi \mu_0 f_0 \sigma_c)^{-1/2} \quad (6)$$

the skin depth,  $\delta$ , which equals  $2.1 \mu\text{m}$  at 1 GHz for Cu with  $\sigma_c = 5.8 \times 10^7 \Omega^{-1}\text{m}^{-1}$ .

The intrinsic impedance  $Z$  (often called “ $R$  over  $Q$ ”) of a cavity,

$$Z = \frac{R}{Q} = \frac{V_0^2}{2\omega W} \quad (7)$$

(but often twice this!), contains the free-space impedance  $377 \Omega$ , the transit time factor included in  $V_0$  and a measure for radial concentration of stored energy. The  $E_{010}$  pillbox with  $\ell = \lambda_0/2$  has  $Z = 121 \Omega$ . Shape (b) may yield 10 to 20% more (e.g. LEP [3, 4] with  $Z = 138 \Omega$ ). For higher modes the entity “loss factor”  $= \omega Z/2$  (an inverse capacitance) is often used.

Non-axisymmetric structures produce rf quadrupole focussing. Inverse focal length of a pillbox with flat (slit or flat oval) apertures,  $V_0 \cos \omega t$  and particle energy  $eU/\text{electron-volt}$ .

$$F^{-1} = \pm \frac{\pi V_0}{\lambda_0 U} \sin(\omega t) \quad (8)$$

Other shapes add numerical factors  $\leq 1$ . The shape of the beam apertures (however small!) governs focussing as much as the cavity shape.

Strings of coupled cavities (cells) are used for large total voltages. The period is nearly always made  $\lambda_0/2$  for  $v = c$ . Direct intercell coupling (electric  $\pi$ -mode) is most convenient. The closeness of  $N - 1$  unwanted modes near  $f_\pi$  limits the number  $N$  of cells, but very successful designs with  $N = 5$  or 7 and a central power coupler exist (with one piston servo-tuner on either side for common tune and balance). Re-entrant cavities require magnetic coupling via off-centre apertures in the separating walls, Fig. 1(b); typical bandwidth  $(f_0 - f_\pi)/f_\pi = 1.3$  [3]. More about passband properties of multicell structures can be found in [4, 5].

Longer strings are coupled by means of parallel side-coupling to a coaxial manifold [6] or by

resonant (bi-periodic) coupling (e.g. [7]): a flat  $E_{010}$  coupling cell is inserted between adjacent accelerating cells, Fig.1(c), so as to form an electrical  $\pi/2$  mode with vanishing field in the coupling cells.

Mechanical construction is mostly from forged discs and rolled cylinders, shape (b), or from thick sheet metal, generally OFHC copper, by electron-beam welding. Exposing such welds to brazing temperatures (for subsequent addition of auxiliaries) has led to problems and should be avoided. Aluminium has been used successfully [8] but is inferior in electrical conductivity, heat conductivity and thermal expansion; it requires surface treatment to reduce secondary emission. Copper plating on steel has also been successful [7]. Servo-tuners in the form of pistons protruding radially inward from the cavity walls compensate thermal expansion and beam loading. Finger contacts to the cavity wall tend to be required and are critical components.

Power transmission from the klystron is usually converted from waveguide to coaxial close to the cavity to accommodate an  $\text{Al}_2\text{O}_3$  vacuum window (disc or cylinder) and a loop coupler. Direct waveguide coupling to the cavity has been used but is not recommended at the UHF frequencies employed here (large window exposed to cavity fields and beam). C.w. powers per coupler substantially above 200 kW are considered high but up to 1 MW has been reached and is routine in klystron outputs.

Fundamental-frequency beam loading is quantitatively treated in [9]. It produces a reduction in effective  $Q$  and detuning (bunches passing off-peak for phase stability). While much less drastic than in SC cavities it may require circulator-protection of the klystron power-source from reflected power.

Higher-mode dampers might not be required (no effect on hardware as with superconductivity, multi-bunch instability possibly avoided due to natural spread of resonances, no effect on single-bunch instability for which a large aperture is the only remedy), while very high beam current and close bunch spacing may require such damping in every cell. Discrimination from  $E_{010}$  by a combination of coupler orientation and choice of high-pass cut-off in the line leading to power absorption. In single cells with large apertures the adjacent vacuum chamber may play part of that role. A radical solution is the choke-mode cavity [10]:

a complete azimuthal gap in the cavity wall communicates with a resistive load via a radial transmission line. A notch filter (a circular groove in the line) provides a short circuit of the gap at  $f_0$ .

Large H-mode storage cavities ( $Q > 10^5$ ) may be side-coupled to the accelerating structure in order to increase the stored energy – so as to alleviate transient beam loading – or to increase the effective  $Q$  in case of a very large bunch spacing. In the first case [11] an  $H_{013}$  cylindrical storage cavity ( $Q = 2 \times 10^5$  at 500 MHz) is coupled to an accelerating cavity via an intermediate (dummy) cavity. The second case [3] is peculiar to the large bunch spacing of LEP: a spherical storage cavity ( $H_{011}$ ,  $Q = 1.6 \times 10^5$  at 350 MHz) is directly coupled to the central cell of a 5-cell accelerating structure. Both modes of the coupled system (storage cavity versus five-cell structure) are driven in such a way as to make the stored energy oscillate between the two systems, spending half the time, on average, in the low-loss storage cavity.

## References

- [1] H.R.L. Lamont, Wave Guides, Methuen & Co. Ltd., Wiley (1959)
- [2] The MAFIA Collaboration, Darmstadt, Germany; M. Bastsch et al, Int. Conf. on Electromagnetic Field Problems and Applications (1992) p.405
- [3] P. Brown et al, PAC89, p.1128
- [4] I. Wilson, H. Henke, CERN Report 89-09 (1989)
- [5] G. Dôme, in Linear Accelerators, North-Holland (1970) p.637
- [6] R.M. Sundelin et al, PAC77, p.1686
- [7] T. Higo et al, PAC87, p.1945
- [8] M.A. Allen et al, PAC 77, p.1780
- [9] P.B. Wilson, Proc. IX Int. Conf. on High Energy Accel. (1974), p.57
- [10] T. Shintake, PAC93, p.1048
- [11] Y. Yamazaki, T. Kageyama, PA 44 (1994) 107

### 7.3.8.2 Single cell cavities

*R. Rimmer, JLab*

For high charge per bunch and high average current machines, the severe requirement on HOM damping dictates the use of single cell cavities.  $Q$ 's of the HOM's range typically from a few hundred to a few thousand. See Figs.2 and 3, and Tab.1. An updated and simplified version of this cavity was developed for use in damping rings or future light sources [6].

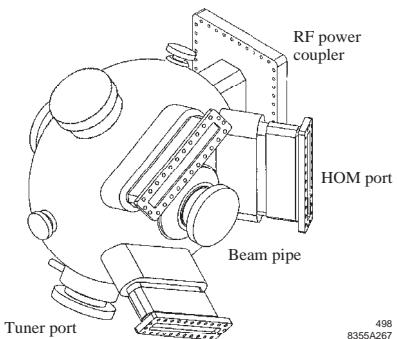


Figure 2: Normal conducting, reentrant cavity.

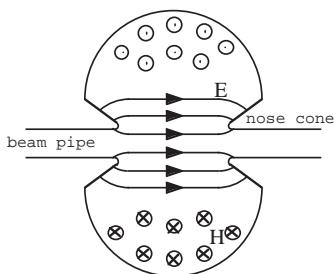
Figure 3: Field pattern of cavity in Fig.2 showing nose cones for  $R/Q$  increase.

Table 1 shows parameters for the PEP-II LER cavity of Fig.2 [1]

|  |   |
|--|---|
| $f$ [MHz]                                  | 476   |
| $R/Q$ [ $\Omega$ ] [2]                     | 116   |
| $Q_o$ calc. 2D                             | 45,000  |
| $\Delta Q_{HOM}$ ports [%]                 | -10   |
| $\Delta Q_{other\ ports, surf.\ finish}$   | -13   |
| $\Delta Q_{temp\ rise\ at\ 150\ kW}$       | -10   |
| $Q_{operating\ power}$                     | $\sim 30,000$                                 |
| $P_{wall}$ [kW] [3]                        | 103   |
| $P_{wall}/Area$ [W/cm <sup>2</sup> ]       | 15-23   |
| max $P_{wall}/Area$ [W/cm <sup>2</sup> ]   | 80  |
| Reynolds No.                               | 50,000  |
| film coeff. [W/cm <sup>2</sup> °C]*        | 1.8   |
| $R_{shunt}$ ( $V^2/2P$ ) [ $M\Omega$ ] [2] | 3.5   |
| $V_{cav}$ [MV]                             | 0.85  |
| $E_{acc}$ [MV/m]                           | 3.9   |
| $E_{surf,\ max}$ nose [MV/m]               | 23  |
| $I_{beam}$ [A]                             | 2.25  |
| $P_{beam}$ [kW]                            | 308   |
| $P_{window, full\ beam}$ [kW]              | 411   |
| $R_{HOM, worst}$ (long.) [ $k\Omega$ ] [5] | $1800^{\dagger\dagger} (\sim 2)^{\dagger}$    |
| $R_{HOM, worst}$ (transv.) [ $k\Omega/m$ ] | $31000^{\dagger\dagger} (\sim 160)^{\dagger}$ |

\* see [4] for more on cooling, †† undamped, ‡‡ damped.

**Tuner** For normal conducting cavities the most common type of movable tuner in the piston type as shown in Fig.4 [7].

Typical dimensions at UHF would be 9 cm diameter, 5 cm travel, 4 mm gap between plunger and wall, rhodium plated over finger region, Glidcop™ alloy 25 spring contact fingers. Depth of fingers chosen to be non-resonant at rf frequency or harmonics. Tuning range -400 to +800 kHz.

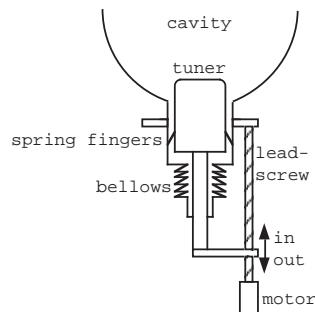


Figure 4: Piston tuning plunger.

**Coupler** Coupling is generally by a loop or aperture as shown in Figs.5 and 6 [1]. Surfaces exposed to rf currents must be well cooled. Loop couplers may be rotatable to allow adjustment of the coupling factor. Windows are often derived from those used on high-power klystrons and are typically derated by 50% because of the less well controlled vacuum conditions. A coating such as titanium nitride is usually applied to the vacuum side of the ceramic to prevent multipactor and surface charging. Designs for 300-500 kW c.w. are available [8, 9, 10].

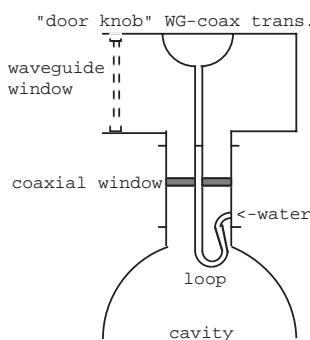


Figure 5: Loop coupler.

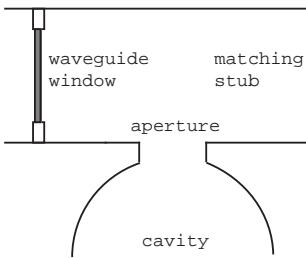


Figure 6: Aperture coupling.

## References

- [1] PEP-II conceptual design report (1993), LBL-PUB-5379, SLAC-418, CALT-68-1869, UCRL-ID-114055, UC-IIRPA-93-01
- [2] R. Rimmer et al, PAC 91, p.819
- [3] R. Rimmer et al, EPAC 94, p.2101
- [4] I. Wilson, CERN Acc. School (1991), CERN 92-03, Vol.II, p.391
- [5] R.A. Rimmer et al, EPAC 96, p.2035; R.A. Rimmer, J.M. Byrd, D. Li, PRST-AB 3 (2000) 102001
- [6] R.A. Rimmer et al, PAC 01; [www-library.lbl.gov/docs/LBNL/479/69/PDF/LBNL-47969.pdf](http://www-library.lbl.gov/docs/LBNL/479/69/PDF/LBNL-47969.pdf)
- [7] H.D. Schwarz et al, PAC 97, p.3039
- [8] M. Neubauer et al, EPAC 96, p.2059
- [9] M. Neubauer et al, PAC 95, p.1803
- [10] E. Chojnacki et al, PAC 97, p.3177

### 7.3.9 Superconducting Cavities for $v_p = c$ Linacs, Storage Rings, & Synchrotrons *D. Proch, DESY* *Updated by H. Padamsee, Cornell*

Superconducting (SC) structures from 500 MHz to 3.9 GHz for either pulsed or c.w. operation are now in operation for a variety of accelerators. Operating temperatures range from 4.2 K (low frequency) to 1.8 K (high frequency). As of 2010, there are 500 accumulated meters of SC structures in operating accelerators. An additional 700 meters are under preparation for approved projects, new accelerators and upgrades. The gradients in operation for linacs have risen steadily from 5-10 MV/m in the 1990's to 15-20 MV/m in 2010. Gradients continue to rise toward 35 MV/m for the next round of demanding projects. Beam currents in c.w. storage rings have risen from 100 mA to 1.5 A. Books and review articles surveying the subject are [1, 2, 3] with many relevant references.

The fundamental advantage of SC over normal conducting (NC) cavities is the low rf wall loss. While typical  $Q$ 's for NC are  $10^4$ , for SC they can be  $10^9$  to  $10^{10}$  giving an rf power loss advantage of  $10^5$  to  $10^6$ . After taking into account the need for refrigeration, generally inefficient, there is a substantial net ac wall plug power saving as shown in the c.w. storage ring application example of Tab.1.

Table 1: Comparison of power consumptions of NC and SC Structures at 500 MHz, 1 MV/m accelerating gradient with 1 m length.

| value                         | SC             | NC             |
|-------------------------------|----------------|----------------|
| R/Q [ $\Omega$ ] <sup>*</sup> | 185            | 370            |
| Q                             | $4 \cdot 10^9$ | $4 \cdot 10^4$ |
| P(4.2K) [W]                   | 0.7            | -              |
| P(300K) [KW]                  | 0.35           | 35             |

\* the definition of R used here is  $V^2/2P$

In the SC case, the high  $Q$  leads to low rf losses allowing the cavity design to be optimized for properties other than the shunt impedance, such as: (i) large iris diameter to reduce wakefields; (ii) choice of low operating frequency for low wakes and high stored energy with low beam loading; (iii) reduction of peak electric and magnetic surface fields to reduce field emission and thermal instabilities; (iv) choice of cavity geometry to reduce multipactoring.

SC cavities need only moderate rf power to establish the accelerating gradient. Most of the rf power supplied is for the beam power. For pulsed operation, the low wall loss allows longer pulse length for longer bunch trains, allowing higher conversion efficiency from AC wall-plug power to beam power.

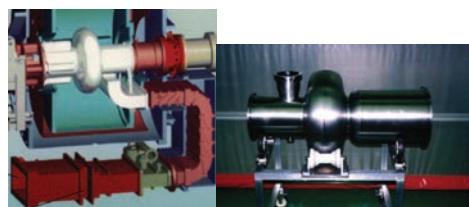


Figure 1: (Left): Single cell 500 MHz cavity [4] for CESR with waveguide input coupler and fluted beam tube on one side to remove HOMs. Single cell 508 MHz cavity for KEK-B with coaxial input coupler port and large beam pipe on one side for propagation of HOMs [5].

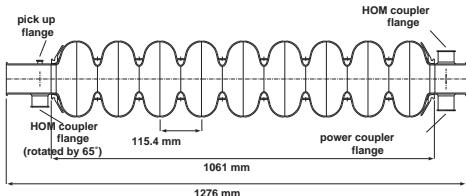


Figure 2: Layout of the components for the 9-cell TESLA-style 1.3 GHz structure [6].

Fig.1 shows the single cell CESR [4] and KEK-B SC cavities [5]. These provide the needed voltage at 5-10 MV/m, with low fundamental mode impedance and strong damping for higher order modes (HOM). Almost all the available rf power is transferred to the beam. The high bunch charge, short bunch length, and tight bunch spacing (10-100 ns) create strong wakefields, higher mode power and multibunch instabilities. With a small number of cells and open beam-hole designs, SC cavities reduce beam-cavity interaction and multibunch instabilities. The low fundamental mode  $R/Q$  and the high c.w. cell voltage reduce transient beam loading and the need for rapid detuning as for example when there are gaps in the bunch train, such as abort or injection gaps.

A nine-cell structure designed for high gradient linacs, such as FLASH at DESY is shown in Fig.2 [6]. The structure is also intended for use in the European X-ray Free Electron Laser (XFEL), and remains a strong candidate for the International Linear Collider (ILC).

The following comparison illustrates the general SC benefits. For a SC linear collider at about 30 MV/m gradient, the peak rf power needed is less than 500 kW/m (almost entirely the beam power) instead of 100 MW/m for NC choices (almost entirely to establish the field) at loaded gradients of 75 MV/m. The pulse length can be ms instead of  $\mu$ s, and the rf frequency  $\sim$ 1 GHz instead of  $\sim$ 10 GHz.

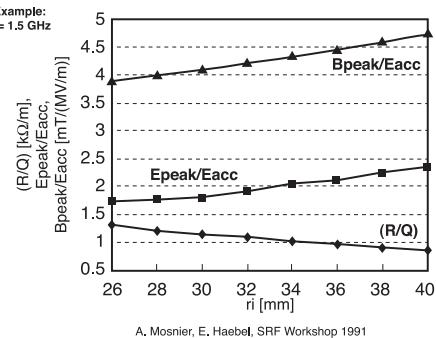
The main disadvantage of SC for linacs is the gradient limit when the surface magnetic field approaches the rf critical field. Present theoretical estimates for Nb rf critical field are 240 mT at 0 K, dropping to 225 mT at 2 K. The highest surface magnetic field obtained in a Nb cavity is near 210 mT, corresponding to an accelerating field of 57 MV/m for the single cell cavity. For the 9-cell cavity of Fig.2 the highest accelerating field achieved is 43 MV/m corresponding to 185 mT.

Currently, high gradient operating accelerators are limited to 25 MV/m by thermal instabilities at local surface imperfections, and field emission. In the former case, the temperature outside a local “defect” rises above  $T_c$  for Nb which results in a quench due to the surrounding Nb entering the normal state. A low thermal conductivity of the material aggravates this condition. In the case of field emission, electrons are accelerated in the rf field and impact the SC surfaces, raising local temperatures and thus lowering  $Q$  and loading the cryogenic system. Quenches from this mechanism are also possible.

**Design considerations** The main figures of merit for a SC accelerating structure are: rf frequency, operating temperature, accelerating voltage ( $V_c$ ), accelerating field ( $E_{acc}$ ), peak surface electric field ( $E_{pk}$ ), peak surface magnetic field ( $H_{pk}$ ), surface resistance ( $R_s$ ), dissipated power ( $P_c$ ), stored energy ( $U$ ),  $Q$  value, geometry factor ( $G = QR_s$ ), geometric shunt impedance ( $R/Q$ ), cell-to-cell coupling for multi-cell structures, Lorentz-Force (LF) detuning coefficient, input power required for beam power ( $P_b$ ), coupling strength of input coupler ( $Q_{ext}$ ), higher order mode (HOM) frequencies, HOM shunt impedances and HOM  $Q$  values. Mechanical properties also play a role to ensure stability under atmospheric loading and temperature differentials, to minimize Lorentz-force (LF) detuning, and to keep microphonics detuning under control. Finally, input and output power coupling issues interact with cavity design. Tuners are required to control the rf frequency during operation. In general there are many trade-offs between competing design requirements. For example, a large aperture will improve the propagation of HOMs out of the structure, but will drop the  $R/Q$  and increase the peak surface electric and magnetic fields.

**Frequency** Both SC and microwave properties of the cavity depend upon the rf frequency. High frequencies are characterized by high  $R/Q$ , small surface area with low stored energy, and low dark current. Low frequencies are characterized by low wake fields, both longitudinal and transverse, low HOM power, long structures for a given cell number, and low surface resistance for better thermal stability at high gradient.

**Number of cells** A large number makes for structure economy but entails trapped HOMs, field flatness sensitivity to tuning errors, and calls for high power input per coupler.

Figure 3: Effect of iris diameter on  $H_p/E_{acc}$ .

Cell Shape Most structures are based on the elliptical cavity which emerged from the “spherical” shape first developed to eliminate multipacting. The tilt of the cell wall provides stiffness against mechanical deformations and is a better geometry for acid draining and water rinsing for surface preparation. The cavity shape is designed for low  $E_{pk}/E_{acc}$  to minimize field emission, low  $H_{pk}/E_{acc}$  for best thermal stability, and high cell-to-cell coupling for enhancement of field flatness. Parameters to be manipulated for optimization are cell diameter at equator, iris diameter and shape of transition from equator to iris.

Iris diameter Increasing the iris diameter increases the cell-to-cell coupling and the peak field ratios ( $E_{pk}/E_{acc}$ ,  $H_{pk}/E_{acc}$ ) but decreases the characteristic shunt impedance ( $R/Q$ ) as well as loss factors of the higher modes ( $k_{ll}$  and  $k_t$ ). For the TESLA shape, Fig.3 shows the effect on  $H_{pk}/E_{acc}$ . The other parameters vary linearly over the range shown in Tab.2.

Table 2: Parameter variations with iris radius.

| Parameter                  | $r = 35$ mm | $r = 51$ mm |
|----------------------------|-------------|-------------|
| $E_p/E_{acc}$              | 2           | 2.34        |
| coupling [%]               | 2           | 5.5         |
| $R/Q [\Omega/\text{cell}]$ | 114         | 78          |

The optimum iris diameter depends on operating conditions. For high gradient application the peak field ratios must be minimized (smaller iris). For high current applications, the HOM loss factors must be small (larger iris). The scaling of the loss factors is given by

$$k_{\parallel} \propto 1/(r^2 \sqrt{\sigma}); \quad k_{\perp} \propto \sqrt{\sigma}/r^3 \quad (1)$$

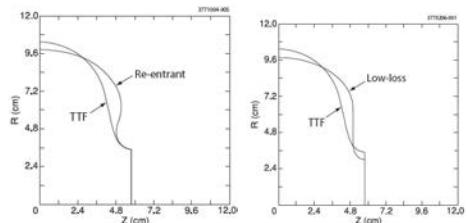
$r$  and  $\sigma$  being iris radius and bunch length respectively.

Table 3: Loss factors for a 9-cell TESLA cavity.

|   |              |
|---|--------------|
| $k_{\parallel}$ for $\sigma_{rms} = 0.7$ mm | 10.2 [V/pC]  |
| $k_{\perp}$ for $\sigma_{rms} = 0.7$ mm     | 15.1 [V/pCm] |

Tab.3 gives loss factors for the 1.3 GHz TESLA structures. Therefore large reduction of the loss factors can be gained at moderate costs of the  $R/Q$  value. This is one primary benefit of SC cavities. The decrease of  $R/Q$  implies increase of rf loss at a given accelerating gradient. Because of the very high  $Q$  of a SC cavity, the additional amount of rf and cryogenic cooling power is acceptable.

Increasing the surface area of the equatorial region, lowers the peak surface magnetic field. Reducing the iris aperture also lowers the peak surface fields, but raises the wakefields. New cavity geometries [7] following these ideas have adopted the “Re-Entrant” ( $\Omega$ -shape) or the “Low-Loss” / ICHIRO shape. Fig.4 compares cell profiles. Table 4 compares important parameters. The new shapes lower  $H_{pk}/E_{acc}$  by 10-15% and raise the shunt impedance ( $R/Q$ ) and the geometry factor ( $G$ ), but raise the surface electric field by 15-20%. The higher value for the product  $G \cdot R/Q$  results in lower rf dissipation (hence “low-loss”). The Re-entrant shape also has higher cell-to-cell coupling.

Figure 4: Comparison of new cell shapes [7] to lower  $H_{pk}/E_{acc}$  with the TESLA-cell shape of Fig.1.

### Mechanical properties (See Sec.5.8)

Lorentz force detuning The interaction of the magnetic surface field and the surface current in a resonator results in a Lorentz force. The EM field exerts a pressure on the cavity wall,

$$P = \frac{1}{4}(\mu_0 H^2 - \epsilon_0 E^2) \quad (2)$$

which results in a deformation of the cells and a change  $\Delta V$  of their volume. The consequence is a frequency shift

$$\frac{\Delta f}{f_0} = \frac{1}{4W} \int_{\Delta V} (\epsilon_0 E^2 - \mu_0 H^2) dV \quad (3)$$

Table 4: Comparison of rf parameters of the new shapes with the TTF shape [7].

| Parameter                                    | Unit                   | TTF   | Re-entrant<br>70 | Low-Loss | Re-entrant<br>60 |
|--|------------------------|-------|------------------|----------|------------------|
| Aperture                                     | mm                     | 70    | 70               | 60       | 60               |
| $H_{pk}/E_{acc}$                             | Oe/MV/m                | 41.5  | 37.8             | 36.1     | 35.4             |
| $E_{pk}/E_{acc}$                             |                        | 1.98  | 2.4              | 2.36     | 2.28             |
| GR/Q   | $\Omega^2$             | 30840 | 33762            | 37970    | 41208            |
| Cell-Cell<br>Coupling<br>Factor              | %                      | 1.9   | 2.38             | 1.52     | 1.57             |
| Loss Factor<br>(long) $\sigma_z =$<br>1 mm   | V/pC                   | 1.46  | 1.45             | 1.72     |                  |
| Loss Factor<br>(transv) $\sigma_z =$<br>1 mm | V/pC/cm <sup>2</sup>   | 0.23  | 0.23             | 0.38     |                  |
| LF Detuning<br>Coefficient                   | Hz/(MV/m) <sup>2</sup> | -0.74 | -0.81            | -0.83    |                  |

where  $W$  is the stored energy and  $f_0$  the resonant frequency of the unperturbed cavity. The frequency shift can be evaluated with mechanical structure codes (e.g. ANSYS) with input of cavity fields from codes such as MAFIA. The net effect is a bending, inward at the iris and outward at the equator. Owing to the high Q, both changes lower the frequency significantly in units of the bandwidth, even though the deflections are only of  $\mu\text{m}$  size. In the TESLA-style cavity, the iris area is stiffened with rings between adjacent cells as shown in Fig.2. Owing to the finite elasticity of the tuner mechanism which fixes the cavity length, there is a net shortening of the cavity, lowering  $f_0$  further. The relevant spring constant is measured to be 37.5 kN/mm. At 25 MV/m (in the TESLA-shape cavity) this contribution is -300 Hz leading to a total detuning of -700 Hz. Were the shape not stiffened, the total detuning would be about -1200 Hz.

**Multi-cell Cavities** A typical SC accelerating structure consists of a chain of coupled cells operating in the SW TM<sub>010</sub> mode, where the phase of the instantaneous electric field in adjacent cells is shifted by  $\pi$  to preserve acceleration as a charged particle traverses each cell in half an rf period. The axial length of each cell equals one half the wavelength of the accelerating wave, thereby establishing  $v_p = c$ . A SW accelerating structure can be analyzed as a chain of coupled resonators. The individual cells are described by single LCR resonators, the coupling between cells might be electric, magnetic or a combination. Electric

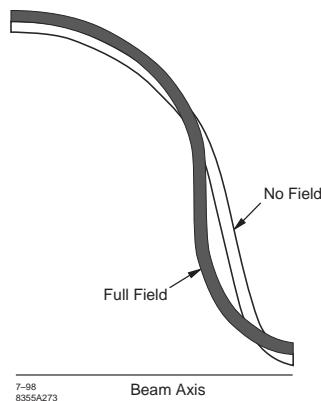


Figure 5: Profile change of un-stiffened cell with Lorentz detuning in effect.

coupling is modeled by coupling capacitor  $C_k$ , the coupling factor  $k$  is the ratio of the cell capacitance to the coupling capacitance. For the TESLA cavity,  $k = 1.8\%$ . The end cells of the accelerating structure are modeled as half cells or full cells. For N cells, N different modes having different resonance frequencies exist. Each mode exhibits a different amount of the stored energy in the individual cells. The amplitude of the accelerating electric field on the axis of the structure is calculated. Electric coupling between adjacent cells through the iris hole is usually dominant in the SC case. The resonance frequencies of the structure are:

$$\omega_q = \omega_0 \sqrt{\left[ 1 - \frac{k}{(1+2k)} \left( 1 + \cos \frac{q\pi}{N} \right) \right]} \quad (4)$$

where  $\omega_q$  is the frequency of mode  $q$ ,  $\omega_0$  is the frequency of an individual resonator and  $k$  is the cell-to-cell coupling, which can be evaluated from (4) by the difference in the frequencies of the highest and lowest frequency modes of the pass-band. The amplitude of the axial electric field in the individual cells is given by

$$A_{n,q} = A \sin \left[ \frac{\pi q(n-0.5)}{N} \right] \quad (5)$$

where  $A_{n,q}$  is the electric field amplitude in cell  $n$  for mode  $q$ ;  $N$  the number of cells;  $q$  the mode number;  $1, 2, \dots, N$ ;  $A$  an amplitude factor. The end cells of an accelerating structure are detuned by the presence of the beam pipe at the outer ends. Therefore the shape of the end cells need correction. In the example of the TESLA 9-cell structure (1.3 GHz), the corrected end cell (without

beam pipe) resonates 3 MHz lower than the middle cells.

Tuning Tuning the fundamental mode field profile is necessary because (i) of enhanced electric or magnetic fields in detuned cells; (ii) the input coupler is in the beam pipe at the end of the structure (iii) the  $R/Q$  of the structure is reduced by unequal excitation. Individual cell errors result in field un-flatness along the beam axis. In the  $\pi$ -mode the voltage sensitivity for this departure is

$$\Delta V_\pi \propto N^2/k \quad (6)$$

depending on cell location, the end cell having the highest and the middle cell the least effect. The field profile is measured, the needed adjustments to the cell frequencies are calculated and the tuning is achieved by deforming the cells with a special tool. Great care must be taken to keep the cavity clean during field profile measurement and tuning.

#### Material properties (See also Sec.6.2)

BCS surface resistance In contrast to dc superconductivity, ac currents produce ohmic loss due to penetration of ac fields into the conductor and acceleration of charge carriers not condensed into Cooper pairs. According to the BCS theory, the surface resistance,  $R_{BCS}$ , in the range  $T \leq T_c/2$  is

$$R_{BCS} = A \frac{\omega^2}{T} \exp \left[ - \left( \frac{\Delta_0}{k_B T_c} \right) \frac{T_c}{T} \right] \quad (7)$$

with  $T$  the absolute temperature,  $T_c$  the SC critical temperature,  $\Delta_0$  the energy gap. The factor  $A$  depends on material properties such as electron mean free path, intrinsic coherence length and intrinsic penetration depth. For dc magnetic fields, The SC State is characterized by critical temperature  $T_c$  and the thermodynamic critical field  $H_c$ . The parabolic temperature dependence of  $H_c$  is

$$H_c(T) = H_c(0)[1 - (T/T_c)^2] \quad (8)$$

The external field,  $H_{ext}$ , is expelled from a superconductor if  $H_{ext} < H_c$  (Type I superconductors). For Type II the external field will partially penetrate for  $H_{ext} > H_{c1}$  (the lower critical field) and will completely penetrate at the upper critical field,  $H_{c2}$ . In rf applications the field varies rapidly in the GHz region. The time period is of the order of the SC  $\rightarrow$  NC phase transition. Under such conditions the predicted limit is the superheating field, as verified for the Pb, In, Sn-In alloys, and for Nb [15].

Residual surface resistance At low temperature the measured surface resistance is larger than  $R_{BCS}$  and is described by a temperature independent residual resistance  $R_{res}$ ,

$$R_{meas} = R_{BCS} + R_{res} \quad (9)$$

Causes of this residual resistance are: (i) Magnetic flux frozen in at cool-down, or after a quench when strong temperature gradients generate thermo-electric currents.

$$R_{res-mag} = R_n (H_{ext}/H_{c2}) \quad (10)$$

where  $R_n$  is the normal state surface resistance. For Nb cavities at 1 GHz  $R_{res-mag} \sim 0.35 \text{ n}\Omega/\text{mGauss}$ , averaged over the cavity area; Magnetic shielding for  $Q$  of  $10^{10}$ , the allowed stray field is 10 mG or less. A cylinder of shielding metal around the helium vessel is usually used. Caution is needed to take care of feedthroughs for power, cryogens, etc.

(ii) Surface contamination contributes a few  $\text{n}\Omega$  typically for Nb in 1 GHz range. (iii) Losses to hydride precipitates.

Impurities influence RRR and thermal conductivity. Bulk Nb is contaminated by interstitial (mostly H, C, N, O) and metallic (mostly Ta) impurities. The resulting RRR,  $\sigma 300\text{K}/\sigma 4.2\text{K}(\text{NC})$ , can be calculated by summation of the individual contributions,

$$RRR_{\text{total}}^{-1} = \sum_{\text{impurities}} \frac{\text{impurity fraction}}{RRR_{\text{impurity}}} \quad (11)$$

Tab.5 gives typical numbers for Nb [8].

Table 5: RRR contributions in units of  $10^4$ .

| RRR       | N    | O    | C    | H    | Ta   |
|-----------|------|------|------|------|------|
| 1 wt. ppm | 0.44 | 0.58 | 0.47 | 0.36 | 111. |
| 1 at. ppm | 2.8  | 3.2  | 3.3  | 18.1 | 58   |

As a rule of thumb, the thermal conductivity at 4.2 K and the RRR are related by

$$\kappa_{4.2K} \approx \frac{1}{4} RRR [\text{W/mK}] \quad (12)$$

The temperature dependence of  $\kappa$  can be fit rather accurately with two parameters, RRR and grain size [9]. The first determines the electron contribution and the second the phonon contribution. Fig.6 shows values of  $\kappa$  vs T for various RRR and grain sizes.

Kapitza resistance of Nb The temperature at the rf surface is higher than that of the LHe at the outer wall due to the flux of heat through the

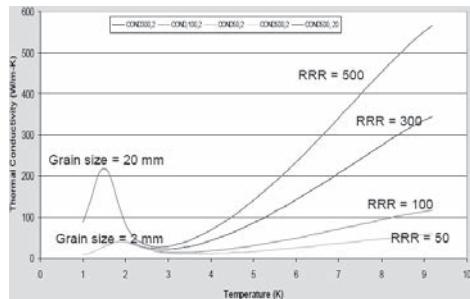


Figure 6: Temperature dependent thermal conductivity of Nb for various RRR and grain size [9].

wall. The temperature difference is governed by the thermal resistance of the Nb and that of the Nb/He interface (Kapitza Resistance). For Nb of RRR 500, the two contributions are about equal. Tab.6 shows the measured Kapitza conductance,  $H_K$  for Nb.

$$\Delta T_{\text{interface}} = \text{power density}/H_K). H_K \sim T^3.$$

Table 6:  $H_K$ , Nb for various conditions (entries are values of  $H_K$  at 1.8 K [W/cm<sup>2</sup>. K]) [10].

| RRR | RRR<br>40 | RRR<br>270 | RRR<br>180 <sup>†</sup> | RRR<br>180 <sup>††</sup> | Ti <sup>♣</sup><br>heat | Ti <sup>♣</sup><br>heat |
|-----|-----------|------------|-------------------------|--------------------------|-------------------------|-------------------------|
| 0.5 | 0.34      | 0.35       | 0.7                     | 0.4                      | 0.85                    |                         |

- ♣ Solid state gettering by Ti for improving RRR, see Nb resonator section in this article
  - ♦ Gettering followed by chemical polish;
- <sup>†</sup> As received; <sup>††</sup> Chemical polished.

**Field limitations** Multipacting Fields in SC cavities can be limited by multipacting (Sec.6.9), by field emission (Sec. 6.13) and by thermal instabilities. With the invention of the round wall (spherical) cavity shape and the elliptical cavity, one-surface MP is no longer a significant problem for velocity-of-light structures. Two point MP does survive near the equator because the electron energies remain between 30-50 eV, which is near the unity cross-over of secondary yield. Conditioning times to reduce the yield can range from fractions of a minute in single cell structures to an hour in multicells.

**Field Emission** Microparticle contaminants are the dominant field emission sources. Increased vigilance in cleanliness during final surface preparation and assembly procedures is

important to keep particulate contamination and associated emission under control. High pressure rinsing and class 100 clean room assembly have achieved high levels of cleanliness in cavity surface preparation, leading to fewer emission sites and major improvements in cavity performance.

**Thermal instability** Thermal instabilities can be simulated by calculating the temperature increase of the inner cavity surface due to the heat flux through the bulk Nb and the Kapitza resistance.

Assuming a defect free surface, the rf heat flux is perpendicular to the wall surface. A thermal instability occurs when  $T_c$  is reached at the inner surface. The instability limit depends on the thermal conductivity of the material as well as on the operating frequency (due to the  $\omega^2$  term in the BCS resistance). Under the realistic assumption of a small NC spot at the rf surface (defect case), the threshold for the thermal instability is lower. It depends on the defect surface resistance as well as on the size of the defect. A high thermal conductivity of the wall material will increase the threshold of the instability.

**Couplers** An input coupler [12] is a device that efficiently transfers power from the rf source to a cavity by providing a good impedance match between the two. The coupler must operate over a wide range of load impedance which varies from a matched at full beam loading to full reflection with no beam. The input power requirement,  $P_f$ , is determined by the operating cavity voltage, beam current and the overhead to compensate peak microphonics detuning, and Lorentz Force detuning [11, 12]:

$$P_f = \frac{V_c^2}{4\frac{R}{Q}Q_{ext}} \left\{ \left[ 1 + \frac{I_b \frac{R}{Q} Q_{ext}}{V_c} \cos \phi_s \right]^2 + \left[ \frac{2Q_{ext}\delta\omega_m}{\omega} \right]^2 \right\} \quad (13)$$

Here  $V_c$  is the cavity voltage,  $I_b$  is the average beam current,  $\phi_s$  is the synchronous phase,  $\delta\omega_m$  is the amplitude of the frequency detuning, and  $\omega$  is the rf frequency. RF feedback loops provide cavity field stability. Environmental microphonic noise creates fluctuations in the cavity resonance frequency and amplitude and phase modulations of the field. This is especially true for high-Q superconducting cavities. The optimum  $Q_{ext}$  is determined by beam loading.

$$Q_{ext} = Vc^2 / [(R/Q)I_b \cos \phi_s] \quad (14)$$

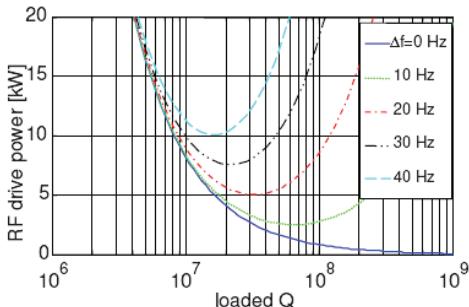


Figure 7: Peak rf drive power as function of  $Q_L$  for a 1.3 GHz, 7-cell cavity at 20 MV/m accelerating gradient [11].

Typical loaded  $Q$  for beam loaded applications range from  $10^5$  to several  $10^6$ . In the case of near zero beam loading, (as for example for an energy recovery linac), the rf power required depends on the microphonics detuning level and choice of loaded  $Q$ , as shown in Fig.7 for 1.3 GHz, 7-cell cavity at 20 MV/m.

The coupler must provide a vacuum barrier between the cavity and the feeder waveguide, allow mechanical flexibility for alignment and thermal contraction during cool down, permit variable coupling strength (external  $Q$ ) in desired cases for different operating modes with beam, and serve as the thermal transition from room temperature to cryogenic temperature with minimal static and dynamic thermal losses. The coupler must be equipped with diagnostic elements (temperature, light detection from vacuum arcs) for safe operation.

There are a variety of waveguide couplers (e.g. Fig.1) well as coaxial couplers that have been developed and reviewed in [12]. Penetration of the cavity wall by couplers should be avoided because of the danger of local field enhancement and multipacting (Sec.6.9). Therefore couplers are placed at the beam pipe. The distance of the coupler port to the first cell determines the range of coupling. Fine adjustment can be done in coaxial design by moving the inner line (a variation of 10 mm results in change of coupling by a factor 10 for the TESLA design coupler). Coupling adjustment for the waveguide coupler can be done by a double stub tuner in the waveguide feed.

RF windows One or two are needed in the coupler line to close the cavity vacuum.

One window design: The window is at room temperature (as for storage rings). The disadvantage of this simple design is that warm surfaces are connected to the cold cavity and that a dust free assembly of the cavity is difficult.

Two window design: In addition to the warm window another window is placed at cold next to the cavity. It enables an early closure of the cavity surface during assembly under clean conditions, but adds complexity and costs.

HOM Couplers [13] The main function is to remove the beam induced power, and to damp the dangerous modes. HOM couplers must also be placed outside the cells to avoid field enhancement. Some HOMs may have very little stored energy in the end-cells due to mode “trapping”. Their suppression becomes difficult. The prime reasons for mode trapping are a small cell-to-cell coupling in the HOM, and a large difference in mode frequency between the end cells and the inner cells.

There are three main types of HOM couplers that are reviewed in [13]: antenna/loop coaxial couplers, waveguide couplers, and beam pipe couplers with absorbers

**Tuners** [14] Slow tuners bring a cavity to resonance, compensating for dimensional changes due to evacuation and cool down, or slow drifts in frequency due to pressure changes in the helium bath. Tuners also compensate the reactive effects of beam loading in high current accelerators to minimize the reflected power. Slow tuners must cover a wide tuning range (of up to several hundred kHz), while providing a resolution of the order of 1 Hz. The whole cavity is changed in length. Slow tuners are usually motor driven. There are a variety of tuner designs reviewed in [14].

Fast tuners compensate static and dynamic Lorentz force detuning, especially at high gradient operation, together with feedforward and feedback. They provide a small tuning range of several cavity bandwidths but with a control bandwidth of several kHz and slew rates of 1  $\mu\text{m}$  in 100  $\mu\text{s}$ . The fast tuner normally has a piezoelectric actuator which is integrated with the mechanical tuner mechanism.

## References

- [1] D. Proch, Rep. Prog. Phys. 61, IOP Pub. (1998), p. 1

- [2] H. Padamsee, J. Knobloch, T. Hays, RF Superconductivity for Accelerators, Wiley (1998)
- [3] H. Padamsee, RF Superconductivity, Science, Technology and Applications, Wiley-VCH, 2009
- [4] H. Padamsee et al, Proc. PAC91, p. 786
- [5] T. Furuya et al, SRF95, p. 729
- [6] B. Aune et al, Superconducting TESLA Cavities, PRST AB 3 (2000) 092001
- [7] R.L. Geng, Physica C 441 (2006) p. 145
- [8] K.K. Schulze, Journal of Metals, No.5 (1981) 33
- [9] F. Koechlin, B. Bonin, Supercond. Sci. Technol. 9, IOP Pub. (1996), p. 453
- [10] A. Boucheffa, M.X. Francois, SRF95, p. 659
- [11] M. Liepe et al, PAC 05, p2642; PAC03, p. 1326
- [12] S. Belomestnykh, Beijing, China, paper WE305 (2007)
- [13] J. Sekutowicz, Proc. LINAC 2006, Linear Accelerator Conference, Knoxville, TN, USA (2006), p. 506
- [14] S. Noguchi, Proc. 13th Workshop on RF Superconductivity, Beijing, China, paper WE303 (2007)
- [15] N. Valles, SRF'11

### 7.3.10 Superconducting Cavities for

$v_p < c$  Linacs

M.P. Kelly, ANL

**Background** Superconducting (SC) TEM cavities, a.k.a ‘drift-tube’ or low-beta cavities, have electromagnetic modes which resemble TEM-modes of a coaxial transmission line. TEM cavities have been operated in ion linacs for particles moving with  $v < c$  for nearly four decades [1]. Most existing TEM cavities are installed in linacs used to accelerate ions to near-Coulomb barrier energies for use in nuclear physics experiments. Cavities are short with dimensions  $l \leq 1$  meter, independently operated and, when combined with modern controls systems, are readily retuned for various ion species or with individual cavities turned off. Generally, superconducting rf cavities have low rf losses, typically  $\sim 10$  Watts per active meter of accelerating structure, or  $\sim 100$  times less than traditional copper structures in c.w. operation. SC TEM-cavity linacs are planned or under study for use in high-energy proton, light- and heavy-ion accelerators for nuclear and high energy physics [2, 3], Accelerator Driven Systems [4], biology and medicine [4], and national defense applications.

Major technical developments in the last decade include the development of new TEM

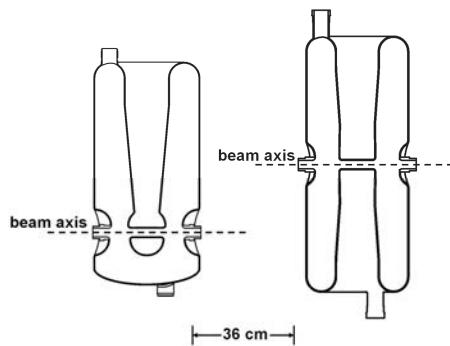


Figure 1: Geometries for quarter- and half-wave cavities.

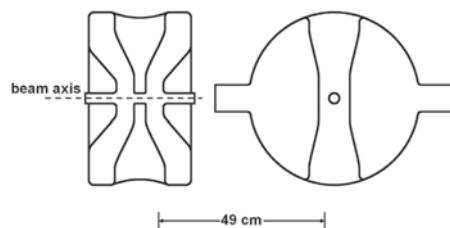


Figure 2: Section views of a single-spoke cavity.

cavity geometries to span the velocity range  $0.01 < \beta < 0.6$  [5]. Together with elliptical-cell structures, SC cavities span the full velocity range up to  $\beta = 1$ . At the same time, a dramatic increase in real estate gradient, from 1 MV/m up to  $\sim 3$  MV/m or more has been due primarily to the use of three dimension simulation tools for cavity design and clean room techniques during fabrication and assembly.

Existing TEM-class cavity types are either quarter-wave ( $\lambda/4$ ) or half-wave structures ( $\lambda/2$ ) [6], as in Figures 1 and 2, operated in the lowest TEM mode with fundamental frequencies ranging from  $\sim 50\text{--}800$  MHz. Physical dimensions are  $0.1 \leq l \leq 1$  meter. Spoke cavities, a type of  $\lambda/2$  structure, have one or more central conductors oriented perpendicularly to a cylindrical outer housing and may be naturally extended to multiple cells per cavity. Other low velocity cavity geometries such as helices, re-entrant cavities and split-rings have been operated in useful linacs, however, performance has been superceded by more recent designs.

To date, TEM-class cavities have been operated near 4 Kelvin since lower temperatures have

| Institution   | Cavity Type     | Frequency (MHz) | Beta (v/c) | # Cavities | Ref.  |
|---------------|-----------------|-----------------|------------|------------|-------|
| Saraf         | HWR             | 176             | 0.09       | 6          | [8]   |
| Triumf        | QWR             | 80, 141         | 0.06-0.11  | 40         | [9]   |
| Mumbai        | QWR             | 150             | 0.1        | 28         | [10]  |
| New Delhi     | QWR             | 97              | 0.08       | 24         | [10]  |
| Canberra      | Split-ring, QWR | 150             | 0.09-0.11  | 14         | [11]  |
| INFN LNL      | QWR             | 80,160          | 0.05-0.13  | 64         | [12]  |
| Kansas State  | Split-ring      | 96,97           | 0.06-0.1   | 14         | [13]* |
| JAERI         | QWR             | 130,260         | 0.1        | 46         | [14]  |
| U. Washington | QWR             | 150             | 0.1-0.2    | 36         | [15]* |
| Florida State | Split-ring      | 97              | 0.07-0.1   | 15         | [16]  |
| Stony Brook   | Split-ring, QWR | 150.4           | 0.07-0.1   | 40         | [17]  |
| Argonne       | Split-ring, QWR | 48, 72, 97,109  | 0.01-0.15  | 64         | [18]  |

offered little performance benefit. It is noted that SC elliptical-cell cavities, used in mostly in electron linacs with velocity  $v/c = \beta \sim 1$ , generally require 2 K operation due to the higher operating frequencies,  $\geq 800$  MHz, and higher associated BCS losses which increase with  $\omega^2$ .

The choice of cavity type for a particular application is best determined based on the required ion species and the associated A/q [7], intensity, energy range and duty cycle. Generally, lower frequencies are beneficial. These benefits include larger accelerating gaps, larger apertures, lower BCS losses, and a reduced longitudinal defocusing factor, with the latter particularly helpful for the efficient transport of high intensity beams. Other considerations, such as the frequency of an RFQ injector or an elliptical-cell cavity or synchrotron post-accelerator may influence the choice of TEM cavities.

The table shows the worlds existing SC linacs based on TEM cavities. All are intended for use relatively with low intensity,  $< 1$  mA, ion beams except that in Ref.[8].

**Design** Electromagnetic designs for TEM-cavities are performed mostly with PC-based three dimensional simulation codes such as MAFIA, Microwave Studio and Analyst. Cavity shapes, and in particular the loading elements, are tailored to minimize peak surface electric and magnetic fields, while at the same time, maintain high shunt impedance. A ratio of peak surface electric fields to accelerating field (accelerating length  $l \equiv n\beta\lambda/2$ , where  $n = \#$  of gaps) of  $E_{PEAK}/E_{ACC} \leq 4$  and of magnetic field,  $B_{PEAK}/B_{ACC} \leq 6$  mT/MV/m is possible for common geometries.

The problem of microphonic or helium pressure induced frequency deviations in TEM cavi-

ties can be largely mitigated in design using FEA simulation tools such as ANSYS or Pro/Engineer. Cavity microphonic bandwidths of  $\sim 10$  Hz or less in operational 4 K linacs are achievable. These microphonics are compensated for using a combination of passive or active tuning systems.

**Fabrication** High-purity bulk niobium ( $RRR \geq 250$ ), available from several vendors, is presently the material of choice for fabrication of both low- and high- $\beta$  SC cavities. It has good SC properties and is formable, machinable and weldable. Fabrication starts with niobium sheet and bar stock and steps include machining, forming, rolling, welding and final surface processing, all having an impact cavity performance. Lead/copper cavities have been built and operated at several laboratories, albeit, with considerably lower performance than for bulk niobium cavities.

Hydroforming or deep-drawing techniques may be used to produce complex niobium shapes and in large quantities if needed. Formed niobium parts are welded together under high vacuum, typically  $2 \times 10^{-5}$  Torr or better, using an electron beam. Demountable joints are less common in new TEM cavities due to the difficulty in producing low loss rf joints. Welded cavities intended for operations are housed inside an integral helium jacket constructed from stainless steel, titanium or niobium.

Generally, TEM cavity fabrication and quality assurance techniques have improved over the past decade leading to improved cavity performance, however, the strict techniques now used to achieve surface magnetic fields with  $B_{peak} > 120$  mT in certain elliptical-cell cavities are not yet uniformly and systematically implemented for TEM cavities.

Preparation of the RF surface using chemistry to remove the  $\sim 100 \mu\text{m}$  layer damaged during fabrication is required to achieve good cavity performance. Electropolishing, the best demonstrated procedure for rf surface preparation for  $v/c = 1$  cavities, has been used to produce many of the best performing TEM cavities [17], though buffered chemical polishing is also used with good results [8].

High-pressure rinsing (HPR) and clean assembly techniques are used at some laboratories as the final step in TEM-cavity rf surface preparation. HPR with filtered deionized water in a class 100 or better clean area effectively removes particulate contamination from the rf surface. Final clean room assembly of the rf power coupler and cavity vacuum hardware has resulted in demonstrations of TEM cavities running with little field emission and with accelerating gradients of  $\sim 10\text{-}20 \text{ MV/m}$  and surface magnetic fields as high as  $\sim 120 \text{ mT}$ .

## References

- [1] L.M. Bollinger, LINAC'98, p.3
- [2] I. Gonin et al, IPAC10 (2010)
- [3] DOE/NSF NSAC, The Frontiers of Nuclear Science (2007)
- [4] DOE Report, Accelerators for Americas Future
- [5] K.W. Shepard et al, PAC05, p.4338
- [6] J. Delayen et al, SRF'03, TuT01
- [7] P.N. Ostroumov et al, PRL 86 (2001) 2798
- [8] A. Nagler et al, LINAC08, p.26
- [9] R.E. Laxdal et al, SRF'09, p.81
- [10] A. Roy, SRF'07, p.24
- [11] N.R. Lobanov et al, SRF'01, p.6
- [12] SPES Technical Design Report 2002, Chapter 5
- [13] K.W. Shepard et al, PAC81, p.3248, \*linac retired from operations
- [14] S. Tacheuchi, SRF87, p.429
- [15] D.W. Storm et al, SRF87, p.173, \*linac retired from operations
- [16] E.G. Myers et al, SRF87, p.405
- [17] J. Sikora et al, SRF 87, p.419
- [18] M.P. Kelly et al, LINAC'08, p.836

### 7.3.11 Superconducting Single Cell Cavities

*S. Belomestnykh, J. Kirchgessner,  
Cornell*

Single cell superconducting cavities are used predominantly in high-intensity storage rings: high-luminosity elementary-particle colliders and

storage-ring-based light sources [1]. They perform different functions: acceleration (fundamental RF systems), bunch lengthening and shortening (beam driven cavities), bunch deflection (crab cavity, see Sec.7.2.10). The existing fundamental RF and crab cavities operate in the frequency range between 350 and 509 MHz. The bunch-lengthening cavities are typically tuned to the third harmonic of the fundamental RF frequency.

In the examples below we will use parameters of the B-cell cavity cryomodule developed for CESR-III [2], listed in Tab.1.

Table 1: Parameters of the CESR B-cell cryomodule.

|  |                 |
|--|-----------------|
| Number of cavities                             | 4               |
| Resonance frequency                            | 499.765 MHz     |
| Accelerating gradient                          | 6 MV/m          |
| Total voltage                                  | 7.2 MV          |
| Fundamental mode $R/Q^*$                       | 89 Ohm          |
| Cavity quality factor $Q_0$                    | $10^9$          |
| Cryostat static heat load                      | 30 W            |
| Cryostat liquid helium volume                  | 520 liters      |
| Beam current                                   | 1 A             |
| Beam power per cavity                          | 325 kW          |
| $Q_{ext}$ of RF input coupler                  | $2 \times 10^5$ |
| Cell loss factor at $\sigma_z = 13 \text{ mm}$ | 0.084 V/pC      |
| Cryomodule loss factor<br>(with one taper)     | 0.48 V/pC       |
| HOM power per cryomodule                       | 13.7 kW         |

\* linac definition

**High current considerations** The advantage of superconducting (SC) cavities over their normal conducting (NC) counterparts in high current storage rings can be outlined as follows [3]. The RF accelerating cells are usually a large contributor to the total storage ring impedance which is a limiting factor to the beam current. In order to keep the impedance as low as possible it is desirable to: (i) achieve as low a broad band and narrow band impedance as possible in each cell, (ii) adequately damp the quality factor  $Q$  of all the higher order resonant modes of the structure, and (iii) use as few accelerating cells as possible.

The use of a SC structure allows one to operate at a higher gradient in each cell because there is no c.w. RF power dissipation limit in cavity walls as in the case of NC cells. Therefore the gradient in each cell can be higher resulting in fewer cells.

The cell can also be shaped entirely differently. The characteristic impedance  $R/Q$  of the

### Sec.7.3: ACCELERATION

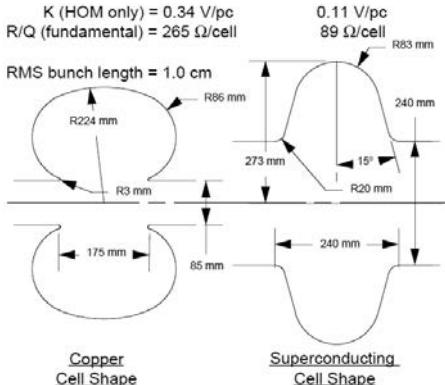


Figure 1: Cavity shape comparison ( $R \equiv V^2/P$  – linac definition), both cells are 500 MHz [4].

cell can be much lower for SC cavities. This permits a large diameter of the beam pipes which allows for the adequate damping of parasitic higher order modes (HOMs) in each cell. The shape of such a cavity has been referred to as the “single mode,” “HOM free,” or “HOM damped” type. A comparison of cell shapes is shown in Fig.1.

The overall advantage in using SC cells is then the product of the gain in gradient and the gain in cell impedance. The required damping of the HOMs is comparable in the two cases.

The other advantages that come with the use of SC cells have to do with the beam loading effects. In accelerators discussed here, the beam loading is mostly reactive and the average effect of this reactive component may be compensated by detuning the cavities by

$$\Delta f/f_{rf} = I(R/Q) \sin \varphi/(2V_c), \quad (1)$$

where  $I$  is the total beam current,  $V_c$  is the cavity voltage and  $\varphi$  is the synchronous phase. With the B-cell cavity parameters

$$\Delta f/f_{rf} = 2.43 \times 10^{-5} \Rightarrow \Delta f = 12.2 \text{ kHz},$$

well below the revolution frequency of CESR, which is 390 kHz, so there should be little danger of exciting a multibunch longitudinal instability due to the fundamental mode impedance.

For proton storage rings which lack synchrotron radiation damping, transient beam loading effects become very important. This happens when the beam current level is changed, as at injection. The transmitter must supply the transient reactive power until the tuner settles to its new equilibrium position to avoid the excitation

of coherent longitudinal oscillations. The minimum power required is  $V\Delta I/4$ , where  $\Delta I$  is the sudden increase in the average beam current.

Any gaps in the beam pulse train, such as clearing gaps, abort gaps or injection gaps, result in a phase modulation of the beam with a maximum excursion (Sec.2.4.3.4)

$$\Delta\varphi_{\max} = (R/Q)(\omega_{rf}/2V_c)I\Delta t, \quad (2)$$

where  $\Delta t$  is the beam gap length. We see that the value for both  $\Delta f$  and  $\Delta\varphi$  are proportional to  $(R/Q)V_c$ , hence there is advantage in lower  $R/Q$  and higher  $V_c$ .

The broadband cavity impedance can be calculated using ABCI [5] or a similar code (Sec.3.2.3). The instability limits can then be calculated by using expressions in Sec.2.4.9. The loss factor  $k_{\parallel}$  (Sec.3.2.6), also calculated by the above mentioned codes, quantifies the energy change of a bunch passing a structure. Fewer number of cell and large diameter beam pipes offered by SC structures result in lower broadband impedance.

Higher order mode (HOM) resonant frequencies of the cells can be excited by the bunched beam, and if the energy level in any of these modes becomes too great, a multibunch instability may occur. The driving term is determined by the magnitude of the beam current with all the Fourier components. The resonant frequency and  $R/Q$  of each mode may be calculated using any of the 2D and 3D frequency-domain eigenvalue solvers, like SUPERFISH [6] or Omega3p [7], and the modes must be damped to some individual value of the quality factor  $Q$  for each mode to prevent HOM field buildup. In the B-cell cavity example cited, for ampere beam currents, the  $Q$  factors of the HOMs must be of the order of 100. The individual  $Q$  limits are determined by programs such as ZAP [8] using the calculated or measured HOM  $R/Q$  values and a machine-specific bunched beam structure.

The broadband and resonant impedance that must be considered includes the entire cavity region, which typically consists of one cavity or a group of cavities, HOM loads, sliding joints, gate valves and has beam line tapers at each end. If a beam tube lined with microwave-absorbing material is used for HOM damping, then a code capable of dealing with such insertions (e.g. CLANS [9]) must be used to calculate  $R/Q$  and  $Q$  values. The loss factor of the cavity region is dominated by that of tapers and HOM loads [10]. The taper

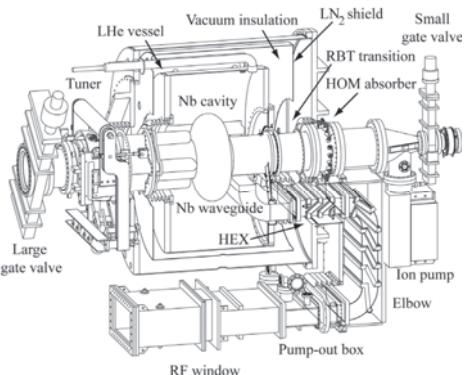


Figure 2: CESR B-cell cryomodule with one taper [2].

shown on the example SC RF cryomodule (Fig.2) has  $k_{||} = 0.169 \text{ V/pC}$  and the pair of ferrite-lined HOM loads has  $k_{||} = 0.288 \text{ V/pC}$  for a 1 cm bunch length.

**Cavities for bunch length manipulation** Having ability to manipulate bunch length may significantly improve performance of high current storage rings. For example, in low to medium energy storage ring light sources, the beam lifetime is usually dominated by the Touschek effect (large-angle intra-beam scattering, Sec.2.4.12). One can improve the Touschek lifetime by increasing momentum acceptance and/or reducing bunch charge density. A particularly attractive option is to lengthen bunches by using harmonic RF cavities [11]. When appropriately phased, relative to fundamental RF, harmonic voltage is applied, it flattens the potential well resulting in lengthened bunches, lower bunch charge density and longer beam lifetime. This is accompanied by an increase in the spread of synchrotron frequencies within the bunch. This bigger synchrotron frequency spread increases damping of the longitudinal coupled-bunch instabilities (so-called Landau damping).

In colliders, it is important for the bunch length to match the beta-function at the interaction point to avoid luminosity decrease by an “hour-glass” effect (Sec.4.1). In this case short bunches are required and can be obtained by adjusting the phase of the harmonic cavity voltage to increase the slope of RF waveform near the synchronous phase [12].

The voltage in the harmonic cavity is generated either by an external generator (active cavity)

or by the beam itself (passive cavity). SC cavities, due to very high Q factor and lower R/Q, have certain advantages over NC cavities in addition to usual advantages of SRF systems. Among those additional advantages are: negligible beam energy loss; operating far from resonance, where the harmonic phase is close to optimum; less sensitivity to an ion cleaning gap. Because of this, the passive cavity operation is often sufficient.

With possible exception of the fundamental power coupler, harmonic cavities must have the same components as their fundamental RF counterparts, and are often scaled versions of the latter [1].

### Power input coupling (see also Sec.6.8)

One can distinguish primary and secondary considerations for the RF power input couplers [13]. The two primary functions are: (i) efficiently transfer power from an RF power source to a beam-loaded cavity; (ii) provide an RF-transparent vacuum barrier, RF window. Among secondary requirements are: provide low heat leak thermal transition from room temperature to the cryogenic environment; confirm to clean cavity assembly procedures; minimize the cavity field perturbation; provide adjustability, when necessary; minimize detrimental effect of multipacting phenomenon.

Waveguide and coaxial antenna type couplers are used for single cell superconducting cavities. The coaxial couplers can have disk or cylindrical ceramic RF vacuum windows. The waveguide couplers employ planar waveguide windows. In choosing, several comparative considerations are germane: size of the waveguide is larger; coupling range of the waveguide is limited to  $Q_L > 10^4$  but can be stronger for the coaxial type; cryostat heat leak is larger in the waveguide case; dissipation in the waveguide walls is less than in the coaxial line and in the absence of an inner conductor waveguide cooling is simpler; coupling adjustment tends to be difficult in both cases but less so for the coaxial case. The achievements in average power coupling to a beam are comparable for both styles of couplers.

The RF vacuum window must handle the entire traveling wave power delivered to the beam as well as vectorially added forward and reflected power at less values of current. The couplers are usually designed for critical coupling at full beam current. Multipacting (Sec.6.9) must be considered in the usual manner for both the window and

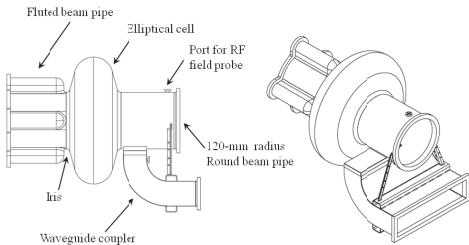


Figure 3: CESR B-cell niobium cavity.

the coupler and thin coating of TiN or TiO are typically used (see Sec.6.15.1).

**HOM damping** Two classes of HOM dampers are used in single-cell designs [14]. The lump-element magnetic or electric couplers (Sec.7.3.9) allow HOM power to propagate and dissipate outside the cryostat, but must have a rejection filter for the fundamental mode power. Several couplers are needed to achieve sufficient damping and sometimes a special design is required to damp the lowest frequency dipole HOMs.

The second class of HOM dampers employs microwave absorbers lining up a large diameter beam pipe just outside the cryomodule at room temperature. This is possible due to the fact that the beam pipe can be made large enough to allow propagation of parasitic modes. In the CESR B-cell case all longitudinal HOMs propagate in the 120-mm radius beam tube, but  $TE_{111}$  and  $TM_{110}$  transverse modes will not propagate. Damping of these modes is accomplished by using “flutes” on one of the beam tubes as shown in Fig.3. The flutes, having only minor effect on the cutoff of longitudinal modes, allow propagation of the two transverse modes of interest to the HOM load. Alternatively, one can use an enlarged beam tube separated from the cell by an iris as it is done in the KEKB SC cavities.

Lossy ferrites are used to absorb HOM power [15] (See Sec.6.5). Ferrite material can be attached either by soldering tiles to metal plates (Elkonite®, a copper-tungsten composite, is used to match thermal expansion of ferrite) or by sintering ferrite powder directly to the beam pipe walls using hot isostatic pressure process [16].

**Cavity tuning** Cavity frequency tuning is accomplished by longitudinally stretching or compressing the cavity along the beam axis [17]. Various alternatives have been employed more or less successfully: at CESR mechanical leverage with

its long stroke and high force, but slow response is used providing the tuning range of 400 kHz with  $df/dz = 225$  kHz/mm; at KEK, fast acting piezoelectric transducers having short stroke and limited force are used; at LEPP both thermal expansion, with its high force, but slow, short stroke response and accompanying heat leak, and magnetostrictive rods with high force, fast response, but short stroke and stray magnetic fields are used.

**Cavity construction** (see also Sec.5.8) SC cavities, at the present state of the technology, can be manufactured of pure niobium or of Cu and then sputter coated on the inside with Nb [18]. The two highest current storage ring single cell superconducting cavities have been of pure Nb. The cost of Nb is not so high to justify all the complication of sputter coating. The gradients that can be achieved with pure Nb are higher than those achieved with Nb sputtered on Cu.

At this time all single cell cavities, with the exception of harmonic ones, operate at 350 to 509 MHz. This is determined by economics, RF transmitter, beam loading and other beam dynamics considerations. The cavity half cells can either be deep drawn or spun. For small quantities, the economic advantage is probably slightly in favor of spinning, but for higher quantities the cost and reproducibility considerations would both favor deep drawing. Cavity welding, completely electron beam welding at this time, for both pure Nb and Nb on Cu construction, can either be accomplished by welding from the inside or by so-called “smooth underbead” external weld. These types of welds are accomplished with techniques that have been developed to lower the energy density of the weld spot by either defocusing or “rastering” of the welding beam. Very smooth, defect-free weld seams inside the cavity are required to avoid the trapping of chemicals and to avoid any thermal magnetic defects that cause breakdown of the superconductivity.

Some provisions must be made to monitor the magnitude of the electric field in the cell. This is necessary to facilitate the voltage control required, and must be done in such a way as not to lower the allowable gradient.

After fabrication the cavity must be chemically etched with 1:1:2 buffered chemical polish to remove 100  $\mu\text{m}$  of material with the acid mixture at a temperature held below 18°C in order to avoid excessive hydrogen going into solution in the Nb. After thorough clean water rinsing,

preferably with high pressure, the cavity must be assembled in a class 100 or better clean room in order to avoid field emission at low field levels.

**Cryostats** The requirements to the cryostat are in general: (i) to hold the cavity in a stable location, (ii) to allow for a cavity tuner and input coupler, and (iii) use as little of liquid helium (LHe) as possible [3]. The CESR B-cell cavity cryostat, complete with the usual liquid nitrogen ( $\text{LN}_2$ ) shield and other required components, is shown in Fig.2. Contributing to the heat leak are conduction and thermal radiation down both beam tubes and waveguide as well as power dissipation, primarily in the waveguide. The large beam tubes, necessary for the transmission of the HOMs, increase both the thermal conduction and radiation. To limit conduction to the helium,  $\text{LN}_2$  cooled heat sink stations appear on each beam tube. Fortunately the large diameter of the tubes allows a significant portion of the thermal radiation to pass through and out the opposite beam tube. The 304 stainless steel beam tubes are both plated with 150  $\mu\text{m}$  of electroless copper to limit the HOM power dissipation in these components. The waveguide components have an additional dissipation load due to RF power transmission so all components are electroless copper plated (250  $\mu\text{m}$ ). A helium gas heat exchanger (HEX) is utilized in the waveguide between 4 K and 77 K. The U-shaped elbow is held at 77 K and the transition to room temperature is made as short as possible. All these steps limit the 4 K heat leak of the cryomodule to 30-35 W. The  $\text{LN}_2$  heat load is about 300 W.

**Other components** Typically both the helium level and the pressure in the LHe vessel must be regulated with control valves between the helium refrigerator and the cryostat. The pressure inside the cryostat must be regulated in order to avoid detuning the cavity due to the finite stiffness of the cell. The CESR B-cell cryomodule exhibits a  $df/dP$  of 250 kHz/bar, a high value primarily caused by the large helium vessel bellows on the fluted beam tube end (Fig.2). With the cavity bandwidth of 2.5 kHz as determined by the coupler, the required pressure regulation is  $\pm 1$  mbar to keep within 1/10 bandwidth. An optimal design of the cryostat bellows could relax this requirement considerably.

Each cavity requires a beam line sliding joint to accommodate the mm-scale motion required for tuning [19]. Also required are isolating

gate valves and sliding joints for installation and changing of a cryomodule. Due to rather high loss factor of the tapers, it is advantageous to keep the beam tube diameter large for an entire group of cavities, with tapers only at the ends rather than a pair of tapers for each cavity. The tapers most likely will have to have water cooled synchrotron radiation masks to absorb, in one place, all the synchrotron radiation power that would have fallen along the cavity region.

## References

- [1] H. Padamsee, RF Superconductivity: Science, Technology, and Applications, WILEY-VCH, 2009, Chapter 11 and references therein
- [2] S. Belomestnykh, PAC99, p.272 & p.980
- [3] H. Padamsee, Jens Knobloch, Tom Hays, RF Superconductivity for Accelerators, John Wiley & Sons, 1998, Chapter 20 and references therein
- [4] J. Kirchgessner, PAC95, p.1469
- [5] Y. H. Chin, PAC93, p.3414
- [6] K. Halbach, R.F. Holsinger, PA 7 (1976) 213
- [7] L.-Q. Lee et al, SLAC-PUB-13529 (2009)
- [8] M. Zisman et al, ZAP User Manual, LBL-21270, ESG-15 (1986)
- [9] D.G. Myakishev and V.P. Yakovlev, PAC96, p.2348
- [10] S. Belomestnykh, SRF 990714-08, Cornell University (1999)
- [11] J.M. Byrd, M. Georgsson, PRST AB 4 (2001) 030701
- [12] S. Belomestnykh et al, PAC03, p.1306
- [13] Ibid. [1], Chapter 8 & ibid. [3], Chapter 18 and references therein
- [14] Ibid. [3], Chapter 16 and references therein
- [15] W. Hartung, PhD Thesis, Cornell University (1996)
- [16] T. Tajima, PhD Thesis, KEK report 2000-10
- [17] Ibid. [3], Chapter 19 and references therein
- [18] Ibid. [3], Chapters 6 & 14 and references therein
- [19] J. Kirchgessner et al, SRF93, p.67

### 7.3.12 Millimeter-Wave Linacs

*D. Whittum, Varian*

Electromagnetic accelerators of mm-scale are of interest in advanced accelerator research for the low peak power required at a prescribed gradient, and the more favorable scaling for pulsed-heating at higher gradients. Whereas an S-Band (2-4 GHz) accelerator requires  $\sim 50$  MW to produce a 20 MeV/m gradient, at W-Band (75-110 GHz) the peak power requirement is  $< 100$  kW. Research

### Sec.7.3: ACCELERATION

problems inherent in mm-wave linacs at any gradient are: (1) fabrication and bonding to achieve good tune and quality factor [1]; (2) wakefields; and (3) higher multipole content. Problems of importance at high-gradient are (4) field emission & breakdown; and (5) pulsed-heating. Items (2) and (4) have been discussed in Sec.2.4.1, 3.2, and 6.13. The outlook and problems for high-gradient acceleration in the wavelength range between X-Band and laser wavelengths are summarized in Fig.1. Depicted are gradients achieved to-date, versus wavelength of the accelerating bucket. (In the case of the plasma accelerators, this means the plasma wavelength, not the EM signal wavelength.) Overlaid in Fig.1 the curve labelled “breakdown” is a fit

$$E_{\text{br}} \approx 25 \frac{\text{MV}}{\text{m}} f^{1/2} \left( 1 + 2.7 f^{3/8} \right) \quad (1)$$

to results of studies at S and X Band, extending only over the solid portion of the curve. Also shown are results of Laser Wakefield Accelerator (LWFA) work at Rutherford Appleton Lab (RAL), and Plasma Beat-Wave Accelerator (PBWA) work at UCLA. Shown but not labelled are DESY S-Band, NLC II, VLEPP, KEK C-Band [2]. The block marked “SLC” extends from 20 MV/m as for a typical structure, to 40 MV/m as for the positron capture section. Finally, in Fig.1 one sees curves quantifying trapping and pulsed heating.

*Trapping* refers to the acceleration of a low-energy ensemble of electrons to relativistic speeds. This mechanism is of concern in high-gradient linacs, where parasitic electrons (e.g. field-emitted electrons) may be accelerated down the beam tube, loading the rf system, and producing unwanted backgrounds for linac instrumentation. The scaling for trapping may be understood from the binding-field expression for acceleration in a sinusoidal wave with  $v_p = c$  [3]. This is an expression relating an electron’s initial phase  $\psi_0$  and speed (normalized to  $c$ )  $\beta_0$ , to its phase  $\psi$ , and speed  $\beta$ , at any later time (see also Sec.4.3.1),

$$\sin \psi = \sin \psi_0 + \frac{1}{\alpha} \left[ \sqrt{\frac{1-\beta}{1+\beta}} - \sqrt{\frac{1-\beta_0}{1+\beta_0}} \right] \quad (2)$$

where the dimensionless quantity  $\alpha$  is

$$\alpha = \frac{e G \lambda}{2 \pi m_e c^2} \quad (3)$$

Particles that are trapped are those following trajectories consistent with  $\beta \rightarrow 1$ , and in the case

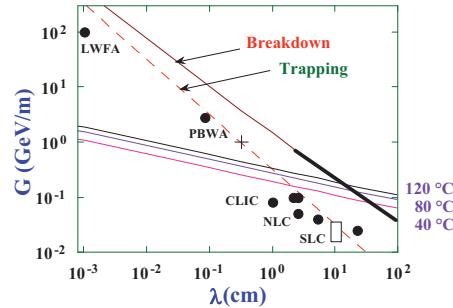


Figure 1: Depicted are gradients achieved to-date, vs wavelength of the accelerating bucket. Overlaid are the 50% trapping condition  $G\lambda \approx 3.2$  MV, an extrapolation of known breakdown scalings, and curves of constant pulsed temperature rise (40°C, 80°C, and 120°C), in one fill-time, in a constant gradient structure with attenuation parameter  $\tau \approx 1$ . We have added a cross-mark as a helpful landmark, corresponding to 1 GeV/m. The corresponding frequency is close to 91.392 GHz (3.28 mm free-space wavelength), the 32nd harmonic of the frequency employed at the SLC, 2.856 GHz.

of a uniform distribution in initial phase, this fraction is

$$f_{\text{trap}} = \frac{1}{2} + \frac{1}{\pi} \sin^{-1} \left( 1 - \frac{1}{\alpha} \sqrt{\frac{1-\beta_0}{1+\beta_0}} \right) \quad (4)$$

For an initially stationary ensemble,  $\beta_0 \ll 1$ , the threshold for trapping corresponds to  $\alpha = 1/2$  or  $G\lambda = \pi m_e c^2/e$ .

*Pulsed heating* refers to heating of the accelerator wall, within a single rf pulse, due to rf dissipation. This phenomenon is described by the diffusion equation for the temperature,  $T$ , within the surface of the conducting accelerator wall,

$$C \frac{\partial T}{\partial t} = \frac{J^2}{\sigma_c} + \kappa \nabla^2 T \quad (5)$$

where the Ohmic dissipation depends on the electrical conductivity  $\sigma_c$ , and the current density  $J$ . This current density appears on the accelerator wall, wherever the magnetic field is non-zero, and evanesces within the surface on the length-scale of the skin-depth. The thermal conductivity is  $\kappa$ , and the heat capacity is  $C$ . Typically the skin-depth  $\ll$  the diffusion depth and the temperature rise at the surface is, for a top-hat pulse,

$$\Delta T = \frac{2S}{\sqrt{\pi \kappa C}} t^{1/2} \quad (6)$$

where  $t$  is the time into the pulse, and  $S$  is the time-averaged dissipated power per unit area at the surface. This temperature rise is of concern for the cyclic stress that results [4]. Operation of the structure as an accelerator requires good tune, and wall  $Q$ ; thus microcracks due to fatigue are of concern as they could disrupt wall currents, lower the  $Q$ , and alter the tune of the structure.

Technologies presently being researched for a W-Band high-gradient linac, with future application to a collider, include: (1) absolute beam position measurement to control wakefield effects; (2) electro-discharge machining for precision manufacture; (3) harmonic acceleration to overcome the length scaling with energy of the final focus chromatic correction section; (4) matrix acceleration to control pulsed heating; (5) beam combining to produce high luminosity at low beam power; (6) high-power microwave switches to match the  $\mu$ s modulator time-scale to the 10-ns fill-time scale for mm-wave linacs; (7) neutral beam collisions to overcome beamstrahlung scalings; and (8) cyclic fatigue limits on temperature rise due to pulsed heating.

Absolute beam position Beam position relative to the null of the dipole wakefields of the accelerating structure [5].

Beam combining Combining of multiple bunches of different energies into one bunch for collisions. This technique permits luminosity scaling quadratically with the number of linac bunches, as opposed to linearly as in the “classical” collider concepts [6].

Electrodischarge machining Cutting by means of spark discharge, between an electrode (wire or stub) placed 50  $\mu$ m from the workpiece immersed in dielectric (oil or water). Rms precision of better than 5  $\mu$ m has been attained on cm-scale structures designed to operate at “32  $\times$  SLAC” or 91.392 GHz [7].

Field-emission Extraction of electrons from a surface in the presence of a large applied voltage, by means of tunneling.

LIGA a German acronym for Lithographie Galvanoformung und Abformung, a sequence of steps corresponding to fabrication of an optical mask, an x-ray mask, x-ray lithography, electroplating, and plastic injection molding. Deep x-ray lithography is of interest for mm-wave structure fabrication, for the mm-depths achievable, the high precision (sub- $\mu$ m), and the possibilities for mass manufacture [8].

Harmonic acceleration Application of multiple frequencies in a linac complex for compensation of correlated energy spread, and reduction of the requirements on, and length of, the chromatic correction section for the collider [6].

Matrix accelerator Generically, an accelerator permitting multiple parallel beams, to reduce exposure time of accelerating structure material to high-intensity fields [9].

Microwave switch In its simplest form, a device with two microwave ports (input and output), and a control that causes the input port to change from a short, to a match transmitting the signal to the output port. At mm-wavelengths, the interest is in fast-switching, within an rf pulse, for purposes of pulse compression. In general, considerable variety is possible in the switch geometry and switch activation technique [10].

Multipole field component The voltage witnessed by a relativistic beam passing through a structure is a harmonic function of transverse coordinate, and as such may be expanded in multipoles. At the same time, this voltage may be decomposed into a sum of modes for frequencies below cutoff in the beam tube, and a broadband component at higher frequencies. The transverse dependence of the accelerating mode will be monopole at lowest order. However in a structure not respecting cylindrical symmetry, higher multipoles may be present in the accelerating mode; thus in a structure respecting inversion symmetry in the two transverse dimensions, the next multipole will be quadrupolar.

Neutral beam collisions Use of combined positron and electron beams in collision to reduce beamstrahlung, pair-production and backgrounds in a high-energy collider [11]. (see Sec.4.14)

Pulsed-heating Heat deposition by Ohmic dissipation of EM energy on a conducting accelerator surface, within a single pulse [12].

W-band The band of frequencies from 75-110 GHz, presently used for vehicular and defense radar, terminal guidance, and remote sensing. Atmospheric absorption in this band can be as low as 0.1 dB/km in dry weather. High-power studies in this wavelength range are presently limited by available power sources. Commercial amplifiers provide less than 5 kW [13]. However, research tubes have provided in excess of 60 kW on the table-top [14], and 1 GW in large-scale experiments [15].

## References

- [1] P.J. Chou et al, Proc. Adv. Acc. Concepts Workshop, AIP Proc. 398 (1997) p.501
- [2] Int. Linear Collider Technical Review Committee Report, G.A. Loew, T. Weiland, eds, (1995) SLAC-R-95-471
- [3] M. Chodorow et al, RSI 26 (1955) 134
- [4] A.M. Freudenthal, J.H. Weiner, JAP 27 (1956) 44
- [5] M. Seidel et al, NIM A404 (1998) 231
- [6] F. Zimmermann, D. H. Whittum, Proc. 2nd Int. Workshop on Electron-Electron Interactions at TeV Energies (IOP, to be published)
- [7] Ron Witherspoon, Inc., 430 Industrial St., Campbell, CA 95008
- [8] A. Nassiri et al, RSI 67 (1996)
- [9] S. Chattopadhyay, D. Whittum, J. Wurtele, Proc. DPF/DPB Summer Study on High Energy Physics, Snowmass (1996) p.356
- [10] S.G. Tantawi et al, AIP Proc. 398 (1997) p.501
- [11] D.H. Whittum, R.H. Siemann, PAC 97
- [12] D. Pritchau et al, PAC 97
- [13] MMW Power Klystrons for Remote Sensing and Communications, Communications and Power Industries (CPI), 45 River Dr., Georgetown, Ontario, Canada L7G2J4
- [14] M. Blank et al, PAC 97
- [15] A.L. Thoop et al, NIM A272 (1988) 15

### 7.3.13 Plasma Accelerators

*E. Esarey, C. Schroeder, LBNL*

A plasma accelerator [1] is a device in which a drive laser or particle beam excites a plasma wave (wakefield) that accelerates a trailing beam. A schematic of a laser-plasma accelerator (LPA), consisting of a drive laser, an electron injector, and a plasma channel that guides the laser pulse and supports the wakefield, is shown in Fig.1.

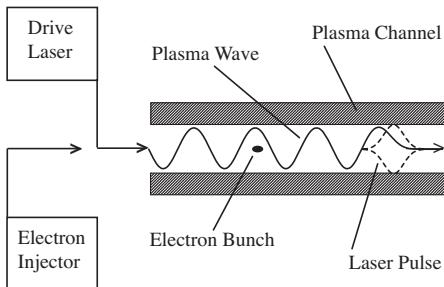


Figure 1: Schematic of an LPA.

An electron plasma wave in the linear regime ( $E_z/E_0 \ll 1$ ) has a electric field  $\vec{E} = -\nabla\phi$  with  $\phi = \hat{\phi}(r) \cos[k_p(z - v_p t)]$ , where  $\omega_p = ck_p = (n_0 e^2 / \epsilon_0 m_e)^{1/2}$  is the electron plasma frequency,  $n_0$  is the ambient plasma density, and  $v_p \simeq c$  is the phase velocity of the wakefield.

A plasma can support a wakefield amplitude on the order of

$$E_0 [\text{V/m}] = cm_e \omega_p / e = 96 n_0^{1/2} [\text{cm}^{-3}] \quad (1)$$

e.g.,  $E_0 \simeq 30 \text{ GV/m}$  for  $n_0 \simeq 10^{17} \text{ cm}^{-3}$ .

**Wavebreaking field** The amplitude of the axial field  $E_z$  of a plasma wave is limited to  $E_z < E_{\text{WB}}$ . Cold, relativistic fluid theory gives [2]

$$E_{\text{WB}} = \sqrt{2}(\gamma_p - 1)^{1/2} E_0 \quad (2)$$

where  $\gamma_p = (1 - v_p^2/c^2)^{-1/2}$ . Thermal effects reduce the wavebreaking limit to [3]

$$(E_{\text{WB}}/E_0)^2 = 2(\gamma_p - 1) - (8/3)\gamma_p(\beta_{\text{th}}\gamma_p)^{1/2} \quad (3)$$

for  $\gamma_p\beta_{\text{th}} < 1$  and  $\gamma_p^2 \gg 1$ , where  $\beta_{\text{th}}^2 = 3T/m_ec^2$  and  $T$  is the electron temperature. Assuming  $\gamma_p\beta_{\text{th}} \gg 1$ ,  $E_{\text{WB}} \approx \beta_{\text{th}}^{-1/2} E_0$ . The wave-breaking amplitude can also be lowered by 2D effects [4] and by density gradients [5].

**Plasma wavelength** The wakefield wavelength in the 1D limit ( $k_p r_\perp \gg 1$ , where  $r_\perp$  is the transverse dimension of the wake) is [6, 7]

$$\lambda_{Np} = \lambda_p \begin{cases} 1 + 3(E_z/E_0)^2/16, & E_z/E_0 \ll 1 \\ (2/\pi)(E_z/E_0), & E_z/E_0 \gg 1 \end{cases} \quad (4)$$

where

$$\lambda_p [\text{m}] = 2\pi/k_p = 3.3 \times 10^4 n_0^{-1/2} [\text{cm}^{-3}] \quad (5)$$

e.g.,  $\lambda_p \simeq 100 \mu\text{m}$  for  $n_0 \simeq 10^{17} \text{ cm}^{-3}$ .

**Phase velocity** The wakefield phase velocity is  $v_p \simeq v_b$  for a particle beam driver with velocity  $v_b$ , and  $v_p \simeq v_g$  for a laser driver with group velocity  $v_g$ . The linear group velocity is [8]

$$\gamma_g \simeq (\lambda^2/\lambda_p^2 + \lambda^2/2\pi^2 r_0^2)^{-1/2} \quad (6)$$

where  $\gamma_g = (1 - v_g^2/c^2)^{-1/2}$ ,  $\lambda$  is the laser wavelength,  $r_0$  is the laser spot size, and  $\lambda \ll \lambda_p, r_0$ .

**Dephasing and tapering** The maximum single-stage energy gain  $W_d \simeq \gamma_d m_e c^2$  of an electron trapped in a 1D plasma wave is limited by dephasing (i.e., the electron outrunning the plasma wave) to [9]

$$\gamma_d \simeq 2\gamma_p^2 \begin{cases} 2E_z/E_0, & 2 \gg (E_z/E_0)^2 \gg 1/4\gamma_p^2 \\ (E_z/E_0)^2, & (E_z/E_0)^2 \gg 2 \end{cases} \quad (7)$$

assuming a non-evolving driver. This occurs over a dephasing length  $L_d$  given roughly by  $W_d = eE_z L_d$ , i.e.,  $L_d \approx \gamma_p^2 \lambda_{Np}$ .

The energy gain can be increased by tapering the plasma density, i.e., increasing the density such that the plasma wavelength decreases to compensate for the slippage between the beam position and the plasma wave phase [10]. In 1D, the density taper to phase-lock a beam (with  $\gamma_b \gg \gamma_p$ ) is [11]

$$n(z)/n_0 = (1 + 3k_p^3 z / 2k_0^2 |\psi_0|)^{-2/3} \quad (8)$$

where  $|\psi_0| = |k_p z(0)|$  is the initial phase of the beam behind the driver.

**Beam loading** The maximum number of beam electrons that can be loaded into a small ( $\ll \lambda_p$ ) axial segment of a wakefield is [12]

$$N_0 \simeq 5 \times 10^5 (E_z/E_0) n_0^{1/2} [\text{cm}^{-3}] A [\text{cm}^2] \quad (9)$$

assuming  $E_z/E_0 < 1$ , where  $A \gg \pi/k_p^2$  is the cross-sectional area of the wake. This beam loading limit is obtained by balancing the wake generated by a short electron bunch with the wake generated by the drive beam. As  $N$  approaches  $N_0$ , the energy spread scales as  $N/N_0$  and the efficiency of converting wake energy to electron energy scales as  $(N/N_0)(2 - N/N_0)$ .

**Plasma wakefield accelerator (PWFA)** An electron beam driver with density  $n_b$  can excite a plasma wave provided that the beam terminates in a time  $\tau_f$  short compared to the plasma period  $\tau_f \omega_p < 1$  [13, 14]. The maximum wakefield amplitude in the linear regime  $n_b/n_0 \ll 1$ , for a drive beam with a Gaussian axial profile of the form  $\rho(\zeta) = n_b \exp(-\zeta^2/2\sigma_z^2)$  with  $\sigma_z$  the axial beam length, is given by [15]

$$E_z/E_0 = \sqrt{2\pi} (n_b/n_0) k_p \sigma_z \exp(-k_p^2 \sigma_z^2/2) H_R \quad (10)$$

assuming  $\gamma_b = (1 - v_b^2/c^2)^{-1/2} \gg 1$ . Here  $H_R$  depends on the radial beam profile. For a flattop radial profile,  $\rho(r) = n_b$  for  $r \leq \sigma_r$  and zero otherwise, where  $\sigma_r$  is the beam radius,

$$H_R = \begin{cases} 1 - k_p \sigma_r K_1(k_p \sigma_r) I_0(k_p r), & \text{for } r < \sigma_r \\ k_p \sigma_r I_1(k_p \sigma_r) K_0(k_p r), & \text{for } r > \sigma_r \end{cases} \quad (11)$$

where  $K_{0,1}$  and  $I_{0,1}$  are modified Bessel functions. For a wide beam  $k_p \sigma_r \gg 1$ ,  $H_R(0) \simeq 1$ . For a narrow beam  $k_p \sigma_r \ll 1$ ,  $H_R(0) \simeq (k_p^2 \sigma_r^2/2)[0.62 - \ln(k_p \sigma_r)]$ .

**Transformer ratio** The ratio of the maximum single-stage energy gain to the initial drive beam energy defines the transformer ratio  $R_t$  in the PWFA. For an axially symmetric drive beam [13],  $R_t \leq 2$ . For a triangular beam [14] with a linear rise over a length  $L_b = N_b \lambda_p$  followed by a rapid termination over a length  $L_f \ll \lambda_p$ ,  $R_t \simeq \pi N_b$ . In the nonlinear regime [16] with  $n_b \simeq n_0/2$ ,  $R_t \simeq \sqrt{2\pi N_b}$ . Using multiple beams may enhance the transformer ratio [13]; with  $M_b$  beams, linear theory predicts  $R_t \leq 2\sqrt{M_b}$ .

**Electron-hose instability** Beams propagating in plasma are subject to instabilities. For example, the number of e-folds for the electron-hose instability is [17, 18]

$$N_e = C_h \gamma_b^{-1/6} k_p z^{1/3} |\zeta|^{2/3} \quad (12)$$

where  $\zeta = z - ct$ , and  $C_h = 3^{3/2}/2^{5/2} \approx 1$  for adiabatic (slow beam rise time) channel formation [17] and  $C_h \lesssim 1$  for non-adiabatic (fast beam rise time) channel formation [18]. Reducing the bunch length is the most effective way to suppress hosing.

**Laser wakefield accelerator (LWFA)** An intense laser pulse in an underdense plasma  $(\lambda/\lambda_p)^2 \ll 1$  can generate a plasma wave via the ponderomotive force  $F_p \propto \nabla a^2$  [19, 20, 21], where  $a^2$  is related to the laser intensity  $I$  by

$$a \simeq 8.5 \times 10^{-10} \lambda [\mu\text{m}] I^{1/2} [\text{W/cm}^2] \quad (13)$$

assuming linear polarization. The amplitude of the quiver momentum of an electron in a laser field is given by  $p_q = m_e c a$ . In the linear regime  $a^2 \ll 1$ , a laser pulse with a normalized intensity profile  $a^2 = \hat{a}^2(r) \exp(-\zeta^2/L^2)$  with  $\zeta = z - ct$  generates a wake of amplitude [20, 21]

$$E_z/E_0 = (\sqrt{\pi}/4) \hat{a}^2(r) k_p L \exp(-k_p^2 L^2/4). \quad (14)$$

For a radial profile  $\hat{a}^2(r) = a_0^2 \exp(-2r^2/r_0^2)$ , the laser power is  $P = 21.5(a_0 r_0 / \lambda)^2$  GW. The wake amplitude is maximum  $E_{\max}/E_0 \simeq 0.38 a_0^2$  when  $L = \lambda_p / \pi \sqrt{2}$ . Figure 2(a) shows the laser excitation of a plasma wave in the mildly non-linear regime.

For a square axial profile of length  $L$  in the 1D limit  $k_p r_0 \gg 1$ , the wakefield amplitude is maximum when  $L = \lambda_{Np}/2$  and is given by [6, 7]

$$E_z/E_0 = (a_0^2/2)(1 + a_0^2/2)^{-1/2}. \quad (15)$$

Note that short pulses  $L < \lambda_p$  are not strongly affected by instabilities.

**Plasma beat wave accelerator (PBWA)** Two long  $L \gg \lambda_p$  laser beams of frequencies  $\omega_{1,2}$  can resonantly drive a plasma wave when  $\omega_1 - \omega_2 \simeq \omega_p$  [19, 22]. As the plasma wave grows, its wavelength increases and becomes detuned from the drive lasers. The maximum plasma wave amplitude at saturation [23] is obtained for  $\omega_1 - \omega_2 = \omega_p$  [ $1 - (9a_1a_2)^{2/3}/8$ ] and is given by

$$E_z/E_0 = 4(a_1a_2/3)^{1/3} \quad (16)$$

assuming  $a_1a_2 \ll 1$ .

**Self-modulated LWFA** A long  $L > \lambda_p$  laser pulse with a power exceeding the self-guiding threshold  $P \geq P_c = 17\lambda_p^2/\lambda^2$  GW can undergo a self-modulation or a forward Raman instability [24, 25, 26]. This results in a strong modulation (with period  $\lambda_p$ ) of the axial intensity profile and the resonant excitation of a large plasma wave. The number of e-folds for self-modulation is [27]

$$N_e \simeq (3^{3/2}/4) [2(P/P_c)k_p|\zeta|(z/Z_R)^2]^{1/3} \quad (17)$$

in the regime  $4[1 - (P/P_c)]^{3/2}/(P/P_c) \ll k_p|\zeta|/(z/Z_R) \ll [2(P/P_c)(k_p/k_0)^3]^{-1}$ , where  $Z_R = \pi r_0^2/\lambda$ . Since  $\lambda_p \sim 1/\sqrt{n_0}$  and  $P_c \sim 1/n_0$ , the conditions  $L > \lambda_p$  and  $P > P_c$  can be met for a sufficiently high plasma density for given laser parameters. Self-trapping of background plasma electrons is possible in this long pulse, high density regime due to heating by the backscattered light and the lower phase velocity.

**Blow-out regime** For a radially bounded drive beam in 3D, the drive intensity can be sufficiently high so as to completely expel all of the plasma electrons from the vicinity of the axis. This high intensity, 3D regime can occur for both laser (referred to as the bubble or cavitation regime) [28, 29] and electron beam (referred to as the blow-out regime) [30] drivers. The blow-out/bubble region of the wake is characterized by an accelerating field that is constant as a function of radius and varies linearly as a function of distance behind the driver, and a focusing field that is linear as a function of radius. Assuming a spherical ion cavity of radius  $r_B$  moving at relativistic velocities, the axial electric field  $E_z$ , radial electric field  $E_r$ , and azimuthal magnetic field  $B_\theta$  within the cavity are [31, 32]

$$\begin{aligned} E_z &\simeq (k_p\zeta/2)E_0, z \\ E_r &\simeq (k_p r/4)E_0, \\ B_\theta &\simeq -(k_p r/4)E_0. \end{aligned}$$

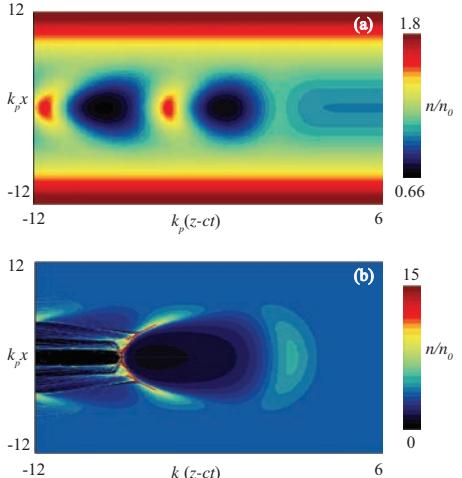


Figure 2: Laser excitation of electron density perturbation  $n/n_0$  [33] in the (a) quasi-linear regime and (b) highly-nonlinear cavitating regime. Profile of the normalized laser vector potential initially has the form  $a = a_0 \exp(-r^2/r_L^2 - z^2/4L^2)$  [propagating to the right and centered at  $k_p(z - ct) = 0$ ], with  $k_p r_L = 5$ , and  $k_p L = 1$  for (a)  $a_0 = 1$  in a matched parabolic plasma channel and (b)  $a_0 = 4$  in a uniform plasma.

The axial electric field is maximum when  $|\zeta| \simeq r_B$ . The radial focusing force on a highly relativistic electron moving along the axis is  $F_r = -e(E_r - B_\theta) = -e(k_p r/2)E_0$ . This regime can have beneficial properties, e.g., because the focusing forces are linear, the emittance of an accelerated electron bunch will be preserved. Positron beams will be defocused in the ion cavity.

For beam drivers the blow-out regime is accessed for parameters such that  $n_b/n_0 > 1$ ,  $k_p \sigma_z < 1$  and  $k_p \sigma_r < 1$ , where  $\sigma_z$  and  $\sigma_r$  are the axial and radial bunch lengths, respectively. For laser drivers, requiring that the radial ponderomotive force be larger than the space charge force of the ion cavity yields the condition  $a_0^2/(1 + a_0^2/2)^{1/2} > k_p^2 r_s^2/2$  to form a cavity. The diameter of the cavity is approximately the nonlinear plasma wavelength,  $\lambda_{Np} \simeq (2/\pi)(E_{max}/E_0)\lambda_p \simeq (2\sqrt{a_0}/\pi)\lambda_p$ , where  $E_{max}$  is the maximum electric field amplitude of the wake and  $a_0 > 1$ . The maximum axial electric field, at  $\zeta \simeq r_B$ , is  $E_m \simeq \sqrt{a_0}E_0$ . Figure 2(b) shows the laser excitation of a plasma wave in the highly nonlinear blowout/bubble regime.

**Multiple drivers** Wakefields may also be excited by multiple drivers. In the linear regime  $E_z/E_0 \ll 1$ , the wakefield produced by a series of  $n$  short  $L = \lambda_p/2$  drivers, separated by an odd multiple of  $\lambda_p/2$ , is  $E_n = nE_1$ , where  $E_1$  is the wake from a single pulse. This holds for a series of laser pulses [34] or electron beams [13]. In the nonlinear regime, resonance with the plasma wave can be maintained by an optimal variation of the pulses lengths and separations. This can result in  $E_n > nE_1$ . For an optimized train of square laser pulses of equal amplitudes [34]

$$E_n/E_0 = (1 + a_0^2/2)^{n/2} - (1 + a_0^2/2)^{-n/2}. \quad (18)$$

Multiple beam drivers (with increasing charge) can be used to greatly increase the transformer ratio beyond the limit for symmetric beams (i.e.,  $R_t > 2$ ) [35]. For  $M_b$  beams with ramped charge relation  $Q_n = (1 - R_{n-1})Q_1$ , where  $n = 2, \dots, M_b$ , the transformer ratio after the  $n^{\text{th}}$  bunch is  $R_n = R_{n-1} + (R_1 - 1)^{n-1}R_1$ .

**Injection techniques** If a small energy spread electron bunch of ultrashort duration ( $\ll \lambda_p$ ) is injected into the wakefield at the proper phase, then the bunch can be accelerated while maintaining a small energy spread. In plasma accelerators, this is challenging for conventional injector technology, since the wakefield wavelength is small (e.g.,  $\lambda_p \simeq 30 \mu\text{m}$  for  $n_0 \simeq 10^{18} \text{ cm}^{-3}$ ), which implies fs timing accuracy.

One method to trigger injection of electrons into a single accelerating bucket (plasma wave period) is to use ultrashort laser pulses. This may be accomplished by either using the ponderomotive force of an additional laser pulse [36], or the beating of two laser pulses to generate a slow phase velocity beat wave [37, 38] (referred to as colliding pulse injection).

Another method to trigger injection is to longitudinally tailor the plasma density [5]. By using a negative plasma density gradient, the local plasma wave phase velocity can be reduced:

$$v_p/c = [1 + (\zeta/k_p)dk_p/dz]^{-1}. \quad (19)$$

Local trapping of electrons in the wake will occur at the point at which the local phase velocity equals the fluid velocity of the plasma electrons. Sharp downward plasma density transitions may also be used for trapping [39].

Injection of background plasma electrons may also be assisted by ionization of electrons from high charge states in the field of the plasma wave. Additional laser pulses may be used for

local ionization and triggered injection [36, 40]. The wakefield driver may also be used for ionization of additional electrons in a high-Z gas and subsequent trapping [41, 42, 43].

**Limits on energy gain in an LPA** In vacuum, the normalized intensity profile of a Gaussian beam will diffract via  $a^2 = (a_0 r_0/r_s)^2 \exp(-2r^2/r_s^2)$  where

$$r_s = r_0(1 + z^2/Z_R^2)^{1/2} \quad (20)$$

is the laser spot size,  $r_0$  is the minimum spot size at the focal point ( $z = 0$ ), and  $Z_R = \pi r_0^2/\lambda$  is the Rayleigh (diffraction) length. Pump depletion (loss of laser energy to the wake) can also limit the propagation length. For an LPA in the limit  $k_p r_0 \gg 1$ , the dephasing length  $L_d$  (see above) is

$$L_d \simeq \frac{\lambda_p^3}{2\lambda^2} \begin{cases} 1, & \text{for } a_0^2 \ll 1, \\ (\sqrt{2}/\pi)a_0/N_p, & \text{for } a_0^2 \gg 1, \end{cases} \quad (21)$$

assuming a non-evolving driver where  $N_p$  is the number of plasma periods behind the drive laser pulse, and the pump depletion length  $L_{pd}$  is [44]

$$L_{pd} \simeq 1.4 (\lambda_p^3/\lambda^2) (2 + a_0^2)/a_0^2. \quad (22)$$

Typically,  $Z_R \ll L_d, L_{pd}$ . If the acceleration length is limited by diffraction, the maximum single-stage energy gain in a LWFA is [1]

$$W_v[\text{GeV}] \simeq 0.7(\lambda/\lambda_p)P[\text{TW}] \quad (23)$$

assuming  $a_0^2 \ll 1$ . Diffraction can be overcome by optical guiding (see below). If the acceleration length is limited by dephasing, the maximum single-stage energy gain in a LWFA is [1]

$$W_d[\text{GeV}] \simeq 0.6I[\text{W/cm}^2]n_0^{-1}[\text{cm}^{-3}] \quad (24)$$

assuming  $a_0^2 \ll 1$  and  $k_p r_0 \gg 1$ . In the nonlinear regime ( $a_0^2 > 1$ ),  $L_d \sim L_{pd}$ , and the the energy gain can be limited by pump depletion, not dephasing, especially if a density taper is used.

Scaling laws for the energy gain in the highly nonlinear blow-out/bubble regime have been obtained through analytical and numerical studies. For example, Gordienko and Pukhov [45] obtain  $\Delta W_{GP}(\text{MeV}) \simeq 0.1(cT_L/\lambda)[P(\text{GW})]^{1/2}$ . Alternatively, Lu et al. [46] find  $\Delta W_L(\text{MeV}) \simeq 0.25(\lambda_p/\lambda)^{4/3}[P(\text{GW})]^{1/3}$ .

**Optical guiding** A long laser pulse  $L > \lambda_p$  can be guided by relativistic self-focusing and/or a preformed plasma channel provided that [47]

$$P > P_c(1 - \Delta n/\Delta n_c) \quad (25)$$

where  $P_c \simeq 17\lambda_p^2/\lambda^2$  GW is the critical power for relativistic self-focusing,  $\Delta n$  is the channel depth,  $\Delta n_c = 1/\pi r_e r_0^2$  is the critical channel depth, or  $\Delta n_c [\text{cm}^{-3}] \simeq 1.1 \times 10^{20}/r_0^2 [\mu\text{m}]$ , where  $r_e = 2.82 \times 10^{-15}$  m, and a parabolic density channel  $n_i = n_0(1 + \Delta n r^2/r_0^2)$  is assumed. For a short pulse  $L < \lambda_p$ , relativistic self-focusing is ineffective [7], and optical guiding requires  $\Delta n \geq \Delta n_c$  [24]. The linear group velocity of a matched laser pulse within a parabolic plasma channel is [8]

$$\beta_g \simeq 1 - \lambda_0^2/2\lambda_{p0}^2 - \lambda^2/2\pi^2 r_0^2 \quad (26)$$

where  $\lambda_{p0}$  is given by the on-axis plasma density.

Plasma channel guiding of short laser pulses was first demonstrated in hydrodynamically formed channels produced by focusing a long pulse laser beam with an axicon lens [48]. In addition to laser-induced channels, guiding has also been demonstrated in plasma channels produced by capillary discharges [49, 50]. Laser-induced channels have typically been limited to a few mm, whereas capillary discharges can be several cm. Preformed channel guiding of relativistically intense  $a_0 \gtrsim 1$  laser pulses [51] was realized using an ignitor-heater method, which is a two laser pulse method for hydrodynamic channel production. In these experiments, an intense ( $>10^{18} \text{ W/cm}^2$ ) laser pulse (500 mJ, 55 fs at input) was propagated through a 2.5 mm channel in a gas jet ( $3 \times 10^{19} \text{ cm}^{-3}$ ) at a spot size of  $\sim 10 \mu\text{m}$ .

**Plasma lens** In addition to acceleration, plasmas can be used for focusing and transport of particle beams [52, 53]. The fields supported in a plasma can provide focusing strengths on the order of MG/cm, several orders of magnitude greater than conventional magnets. Such plasma lens have been considered for use in the final focusing system of future linear colliders [54, 55].

A self-focusing plasma lens can be in two regions determined by the ratio of the plasma density  $n_p$  to the beam density  $n_b$ : the overdense regime  $n_p \gg n_b$  and the underdense regime  $n_p \ll n_b$ . Beam lengths that are long compared to the plasma wavelength are considered. In an overdense plasma lens, the plasma electrons respond to the beam charge and move such that the remaining plasma ions neutralize the space-charge force of the beam. Provided the return current flows mostly outside the beam volume, i.e.,  $k_p r_b \ll 1$ , the remaining beam self-generated magnetic field provides a radial focusing force. For positrons the charge neutralization also occurs by the plasma electron density increasing in

the beam volume. For a linear focusing force,  $F_\perp = -\gamma mc^2 K x_\perp$ , where  $K$  is the focusing strength. The focusing strength for an overdense plasma lens is

$$K(r, \zeta) = \frac{4\pi r_e n_b}{\gamma} G(\zeta) \frac{\partial^2 F(r)}{k_p^2 \partial r^2} \quad (27)$$

where  $G = k_p \int d\zeta' g(\zeta') \sin[k_p(\zeta' - \zeta)]$ , and

$$F = k_p^2 \int dr' r' f(r') I_0(k_p r_<) K_0(k_p r_>), \quad (28)$$

with  $r_<$  ( $r_>$ ) the smaller (larger) of  $r$  and  $r'$ , respectively. The beam density distribution is  $n = n_b f(r) g(\zeta)$ . The focusing force is nonlinear with respect to  $r$  away from the axis. Near the axis,  $K \approx 2\pi r_e n_b / \gamma$ . Any longitudinal variation of the beam density, or excitation of a longitudinal plasma wake, will result in a variation of the focusing force along the beam. Rosenzweig *et al.* first experimentally demonstrated the overdense regime [56]. Ng *et al.* first experimentally studied plasma lens focusing of positrons [57].

If  $k_p r_b > 1$  then the return current flows through the beam canceling the beam current and the self-focusing effect [58, 59]. In this regime, the focusing strength is reduced [58]:

$$K \approx (2\pi r_e n_b / \gamma) [1 + (k_p r_b)^2]^{-1}. \quad (29)$$

In the underdense regime, accessible for  $n_b > n_0/2$  [60, 61], the plasma dynamics are nonlinear. For an electron beam with  $k_p r_b < 1$ , the plasma electrons are expelled from the beam volume, and the immobile ions provide an approximately linear focusing force with

$$K \approx 2\pi r_e n_p / \gamma. \quad (30)$$

For positron beams, the plasma dynamics are different and the electrons are attracted to the positron beam and preform oscillations. The plasma electrons spend a fraction of time within the beam, resulting in a negative charge that provides a focusing force. For a positron beam in the underdense regime, the average focusing force for the beam core is  $K \approx 8r_e n_p / \gamma$  [62].

One limitation to the minimum beam spot size that can be achieved is due to the synchrotron radiation induced during the focusing (referred to as the Oide limit [63]). This limitation can in principle be overcome by employing adiabatic plasma focusing [64], where a thick lens is used and the plasma density is gradually increased (with scale length long compared to the  $\beta$ -function) along the

direction of beam propagation. In adiabatic focusing, the  $\beta$  for high energy particles is always smaller than the matched  $\beta$  for low energy particles, thus the entire beam may be focused even if the beam energy spread increases owing to synchrotron radiation in the plasma lens.

**Particle beam driven experiments** The PWFA mechanism was demonstrated in several experiments. In the pioneering experiments by Rosenzweig *et al.* [65], the wakefield was mapped out by measuring the energy of a witness electron beam that trailed the drive beam by an adjustable delay. Linear and nonlinear wakefields were observed with a maximum energy gain of 0.2 MeV and a maximum accelerating gradient of 5 MV/m.

In 2007, PWFA experiments at SLAC [66] demonstrated the energy doubling of a fraction of electrons in a multi-10 GeV electron bunch using a meter-scale plasma. These experiments used the 50 fs, 42 GeV bunch containing  $1.8 \times 10^{10}$  electrons from the 3 km long SLAC linac, focused to a spot size of  $10 \mu\text{m}$  at the entrance of a 85 cm long column of lithium vapor of density  $2.7 \times 10^{17} \text{ cm}^{-3}$ . In these single-bunch experiments, the front portion of the electron bunch ionized the vapor and generated a large amplitude plasma wakefield in the blow-out regime, which subsequently accelerated a fraction of electrons in the tail of the bunch to energies as high as 85 GeV. The majority of the bunch electrons lost energy, which represents the energy needed to drive the plasma wave. Positron beam driven wakes have also been demonstrated at SLAC [67].

**Laser beam driven experiments** Acceleration of electrons in the PBWA, LWFA, and self-modulated LWFA was demonstrated by several groups, however with large (near 100%) energy spread (see Ref.[1]). For example, the acceleration of self-trapped plasma electrons has been observed in several experiments in the self-modulated LWFA regime. Typically, the laser power ranged from a few to a few tens of TW and the target was a gas jet of a few mm diameter with densities in the range  $10^{19}\text{--}10^{20} \text{ cm}^{-3}$ . The accelerated electron beam charge was large (up to several nC), well-collimated ( $<10$  mrad), and the energy spectrum was typically characterized by an exponential distribution, with the majority of electrons at modest energies (a few MeV) and the number of electrons at high energy (tens to hundreds of MeV) was an exponentially small fraction of the total charge. Although steady progress

was made, prior to 2004, the 100% beam energy spread remained a major limitation.

#### High quality bunches at the 100 MeV level:

In 2004 three groups reported measurement of electron bunches from LPAs with narrow energy spread and a small divergence [68, 69, 70]. The RAL [68] and LOA [69] groups used relatively large laser spot sizes. This effectively increases the diffraction (or Rayleigh range,  $Z_R$ ) of the laser pulse, thereby permitting propagation over distances  $\sim$  the gas jet length. The RAL collaboration used a 0.5 J, 40 fs laser pulse focused ( $25 \mu\text{m}$  spot diameter,  $2.5 \times 10^{18} \text{ W/cm}^2$ ) on a 2 mm gas jet with a plasma density of  $2 \times 10^{18} \text{ cm}^{-3}$ . A narrow energy spread bunch was observed at 78 MeV with 3% FWHM energy spread,  $<5^\circ$  FWHM divergence, and 22 pC of charge [68]. The LOA experiments used a 1 J, 33 fs laser pulse focused ( $21 \mu\text{m}$  FWHM diameter spot,  $3.2 \times 10^{18} \text{ W/cm}^2$ ) on a 3 mm gas jet with a plasma density of  $6 \times 10^{18} \text{ cm}^{-3}$ . A narrow energy spread bunch was observed at 170 MeV with 24% energy spread, 10 mrad FWHM divergence, and 500 pC of charge [69]. The LBNL experiments [70] used a 9 TW, 55 fs laser pulse focused to a relatively tight spot size ( $8.5 \mu\text{m}$  FWHM,  $10^{19} \text{ W/cm}^2$ ). To mitigate the short  $Z_R$  of the laser, a preformed channel (on-axis density  $2 \times 10^{19} \text{ cm}^{-3}$ ) was used to guide the relativistically intense laser pulse through the gas jet. A narrow energy spread bunch was observed at 86 MeV with 3.6 MeV FWHM energy spread, 3 mrad divergence, and  $2 \times 10^9$  electrons. The production of high quality electron bunches in LPAs requires the following four steps. Step 1 is wakefield excitation, which may require the pulse to propagate a sufficiently long distance (e.g., several e-foldings) if relying on self-modulation. Step 2 is trapping and injection of the electrons, which typically occurs when the wake amplitude exceeds some self-trapping threshold. Step 3 is termination of the injection process. If trapping is not terminated, low energy electrons would continuously be injected into the wake over the entire length of acceleration, resulting in a large energy spread. One mechanism to accomplish this is by beam loading, i.e., the injected electron bunch is of sufficient charge so as to reduce the amplitude of the wake below the self-trapping threshold. Step 4 is acceleration of the electron bunch over a distance equal to the dephasing length, such that the trapped bunch exits the

plasma near the top of the separatrix (i.e., the accelerating phase-space bucket), with maximum energy and minimum energy spread.

High quality bunches at the 1 GeV level: In 2006, high quality electron bunches at the 1 GeV level were demonstrated in channel-guided LWFA experiments at LBNL [71]. In these experiments, the energy gain was extended to the GeV range by using higher laser powers (e.g., 40 TW), longer plasma channels (e.g., 3.3 cm), and lower plasma densities (e.g.,  $10^{18} \text{ cm}^{-3}$ ) so as to extend the dephasing length. To overcome the limitations of gas jets, gas-filled capillary discharge waveguides [50] were used to produce cm-scale, lower density plasma channels. A GeV electron bunch ( $\sim 30 \text{ pC}$  charge, percent level energy spread, and 1-2 mrad RMS divergence) was obtained using a 38 TW, 40 fs laser pulse focused to a  $25 \mu\text{m}$  spot radius ( $\sim 10^{18} \text{ W/cm}^2$ ) at the entrance of a  $310 \mu\text{m}$  diameter, 3.3 cm long capillary with an on-axis density of  $4.9 \times 10^{18} \text{ cm}^{-3}$ . Simulations in 2D and 3D confirm that the injection and acceleration mechanism is similar to that which occurs in the gas jet experiments at the 100 MeV level.

High quality bunches from colliding pulses: High quality electron bunches at the 100 MeV level were generated in 2006 by colliding pulse injection at LOA [72]. These experiments used a two pulse, collinear, counter-propagating geometry, in which injection results from the beat wave produced when the backward pulse overlaps the forward drive pulse that generates the wakefield [38]. Specifically, two 30 fs laser pulses with linear polarization and intensities of  $3.4 \times 10^{18} \text{ W/cm}^2$  and  $4.3 \times 10^{17} \text{ W/cm}^2$  were focussed at the edge of a 2 mm supersonic helium gas jet of density  $7.5 \times 10^{18} \text{ cm}^{-3}$ . The addition of the injection pulse triggered production of a high quality electron bunch: a peak energy of  $117 \pm 7 \text{ MeV}$ , an energy spread of  $11 \pm 2\%$ , a charge of  $19 \pm 6.8 \text{ pC}$ , a divergence of  $5.8 \pm 2 \text{ mrad}$ , and a pointing stability of  $0 \pm 1.8 \text{ mrad}$  (for 20 consecutive shots). Further work using colliding pulse injection to control the injected phase space volume has produced 200 MeV electron beams with 1% (FWHM) relative energy spread [73].

High quality bunches from density transitions: Using a density gradient to self-trap electrons from the background plasma, stable electron bunches at the 1 MeV level have been demonstrated by focusing a 10 TW, 50 fs laser

( $2 \times 10^{19} \text{ W/cm}^2$ ) on the downstream edge of gas jet of density  $2 \times 10^{19} \text{ cm}^{-3}$  [74]. Electron bunches were generated (over hundreds of shots) with  $\sim n\text{C}$  charge (15% charge stability), mean momenta 0.76 MeV/c, low absolute momentum spread (170 keV/c FWHM), 20 keV/c momentum stability, 20 keV/c transverse momentum, and 2 mrad (2 keV/c) RMS pointing stability. Other injection experiments using laser-produced density gradients have achieved narrow energy spread beams at the 100 MeV level [75].

## References

- [1] E. Esarey, C. B. Schroeder, W. P. Leemans, Rev. Mod. Phys. 81, 1229 (2009).
- [2] A. I. Akhiezer, R. V. Polovin, Sov. Phys. JETP 3, 696 (1956).
- [3] C. B. Schroeder, E. Esarey, B. A. Shadwick, PR E 72, 055401(R) (2005).
- [4] S. V. Bulanov et al, PRL 78, 4205 (1997).
- [5] S. Bulanov et al, PR E 58, R5257 (1998).
- [6] S. V. Bulanov, V. I. Kirsanov, A. S. Sakharov, JETP Lett. 50, 198 (1989).
- [7] P. Sprangle, E. Esarey, A. Ting, PRL 64, 2011 (1990).
- [8] E. Esarey, W. P. Leemans, PR E 59, 1082 (1999).
- [9] E. Esarey, M. Piloff, Phys. Plasmas 2, 1432 (1995).
- [10] T. Katsouleas, PR A 33, 2056 (1986).
- [11] W. Rittershofer et al, Phys. Plasmas 17, 063104 (2010).
- [12] T. Katsouleas et al., PA 22, 81 (1987).
- [13] R. D. Ruth et al, PA 17, 171 (1985).
- [14] P. Chen et al, PRL 54, 693 (1985).
- [15] W. Lu et al, Phys. Plasmas (2), 063101 (2005).
- [16] J. B. Rosenzweig, PRL 58, 555 (1987).
- [17] D. H. Whittum et al, PRL 67, 991 (1991).
- [18] C. Huang et al, PRL 99, 255001 (2007).
- [19] T. Tajima, J. M. Dawson, PRL 43, 267 (1979).
- [20] L. M. Gorbonov, V. I. Kirsanov, Sov. Phys. JETP 66, 290 (1987).
- [21] P. Sprangle et al, APL 53, 2146 (1988).
- [22] C. Joshi et al, Nature 311, 525 (1984).
- [23] C. M. Tang, P. Sprangle, R. N. Sudan, Phys. Fluids 28, 1974 (1985).
- [24] P. Sprangle et al., PRL 69, 2200 (1992).
- [25] N. E. Andreev et al., JETP Lett., 55 (10), 571–576 (1992).
- [26] T. M. Antonsen, Jr., P. Mora, PRL 69, 2204 (1992).
- [27] E. Esarey et al, PRL 84, 3081 (2000)
- [28] P. Mora, T. M. Antonsen, Jr., PR E 53, R2068 (1996).

- [29] A. Pukhov, J. Meyer-ter-Vehn, *Appl. Phys. B* 74, 355 (2002).
- [30] J. B. Rosenzweig et al, *PR A* 44, R6189 (1991).
- [31] I. Kostyukov, A. Pukhov, S. Kiselev, *Phys. Plasmas* 11, 5256 (2004).
- [32] W. Luet et al, *PRL* 96, 165002 (2006).
- [33] C. B. Schroeder et al, *PRST-AB* 13, 101301 (2010).
- [34] D. Umstadter, E. Esarey, J. Kim, *PRL* 72, 1224 (1994).
- [35] C. Jing et al, *PRL* 98, 144801 (2007).
- [36] D. Umstadter, J. K. Kim, E. Dodd, *PRL* 76, 2073 (1996).
- [37] E. Esarey et al, *PRL* 79, 2682 (1997).
- [38] G. Fibiani et al, *PR E* 70, 016402 (2004).
- [39] H. Suk et al, *PRL* 86, 1011 (2001).
- [40] M. Chen et al., *JAP* 99, 056109 (2006).
- [41] E. Oz et al, *PRL* 98, 084801 (2007).
- [42] A. Pak et al, *PRL* 104, 025003 (2010).
- [43] C. McGuffey et al, *PRL* 104, 025004 (2010).
- [44] B. A. Shadwick, C. B. Schroeder, E. Esarey, *Phys. Plasmas* 16, 056704 (2009).
- [45] S. Gordienko, A. Pukhov, *Phys. Plasmas* 12, 043109 (2005).
- [46] W. Lu et al, *PRST AB* 10, 061301 (2007).
- [47] E. Esarey et al, *IEEE J. Quantum Electron.* 33, 1879 (1997).
- [48] C. G. Durfee III, H. M. Milchberg, *PRL* 71, 2409 (1993).
- [49] Y. Ehrlich et al, *PRL* 77, 4186 (1996).
- [50] A. Butler, D. J. Spence, S. M. Hooker, *PRL* 89, 185003 (2002).
- [51] C. G. R. Geddes et al, *PRL* 95, 145002 (2005).
- [52] P. Chen, PA 20, 171 (1987), P. Chen in *Handbook of Accelerator Physics and Engineering*, World Scientific (2006) p.524
- [53] J. B. Rosenzweig, P. Chen, *PR D* 39, 2039 (1989).
- [54] P. Chen, S. Rajagopalan, J. Rosenzweig, *PR D* 40, 923 (1989).
- [55] P. Chen, C.-K. Ng, S. Rajagopalan, *PR E* 48, 3022 (1993).
- [56] J. B. Rosenzweig et al, *Phys. Fluids B* 2, 1376 (1990).
- [57] J. S. T. Ng et al, *PRL* 87, 244801 (2001).
- [58] D. H. Whittum, A. M. Sessler, PA 34, 89 (1990).
- [59] R. Govil et al, *PRL* 83, 3202 (1999).
- [60] J. J. Su et al, *PR A* 41, 3321 (1990).
- [61] M. C. Thompson et al, *Phys. Plasmas* 17, 073105 (2010).
- [62] P. Chen (unpublished).
- [63] K. Oide, *PRL* 61, 1713 (1988).
- [64] P. Chen et al, *PRL* 64, 1231 (1990).
- [65] J. B. Rosenzweig et al, *PR A* 39, 1586 (1989).
- [66] I. Blumenfeld et al, *Nature* 445, 741 (2007).
- [67] B. E. Blue et al, *PRL* 90, 214801 (2003).
- [68] S. P. D. Mangles et al, *Nature* 431, 535 (2004).
- [69] J. Faure et al, *Nature* 431, 541 (2004).
- [70] C. G. R. Geddes et al, *Nature* 431, 538 (2004).
- [71] W. P. Leemans et al, *Nature Phys.* 2, 696 (2006).
- [72] J. Faure et al, *Nature* 444, 737 (2006).
- [73] C. Rechatin et al, *PRL* 102, 164801 (2009).
- [74] C. G. R. Geddes et al, *PRL* 100, 215004 (2008).
- [75] J. Faure et al, *Phys. Plasmas* 17, 083107 (2010).

## 7.4 BEAM INSTRUMENTATION AND DIAGNOSTICS

### 7.4.1 Composition - Ion Beams

*P. Forck (GSI), P. Strehl (PET)*

For heavy ion accelerators the composition of the accelerated particle beam may be not clear without ambiguity due to the following effects:

- Sputtering ion sources like a Penning ion source (PIG) use an auxiliary gas for plasma generation to produce metal ions by controlled sputtering from a dedicated sputtering electrode made from the required ion. Due to an unfavorable combination between sputtering electrode and the auxiliary gas it may be impossible to separate the auxiliary gas ions, which in general will be produced with high intensity.
- Due to uncontrolled sputtering of materials in an ion source strange ions having nearly the same charge over mass ratio ( $\zeta/A$ ) as the required ion may be produced. In case, separation in the injection area can be very difficult or even not possible and can result in acceleration of wrong ions up to the final energy.
- Using an ECR ion source also ions from the material of the oven may be produced.
- Due to heating up of O-ring sealings by rf-currents, carbon isotopes can be produced which may happen for example in an ECR ion source. Depending on the place of carbon production those ions may also be accelerated.
- In some cases the required ions are generated using gases which can deliver ions of different elements with exactly the same charge

over mass ratio. An example is the generation of carbon ions from CO<sub>2</sub>-gas with an ECR ion source, delivering also ions from all oxygen isotopes.

- Due to leaks in the ion source mainly ions of N<sub>2</sub> and O<sub>2</sub> from the air can contribute to the composition of the accelerated ions.

### Determination of the composition:

- First step should be a check of all available analyzing systems along the machine concerning their resolution and possibilities for separation of strange ions.
- In a second step all isotopes of the materials used in the ion source construction should be listed, including their relative abundance.
- After that all possible charge over mass ratios should be calculated and compared to the  $\zeta/A$ - value of the required ion. Considering the resolution of the analyzing systems it can be decided about a possible separation.
- Typical materials used in ion sources are: W, Ta, Stainless Steel, Cu, Mo, Zr, Al<sub>2</sub>O<sub>3</sub>

In the following the procedure is demonstrated taking a realistic case from heavy ion acceleration requiring <sup>207</sup>Pb<sup>9+</sup> ions. A computer program first calculates the correct  $\zeta/A$ - value (here  $A/\zeta = 206.9759/9 = 22.9973$ ). To constrain the possibilities to reasonable values the following limits have been specified:

- Maximum possible charge number ( $\zeta = 10$ ),
- Minimum mass over charge (22.5),
- Maximum mass over charge number (23.2)

which of course depend on the resolution of the analyzing system and the type of ion source (PIG in this example). Table 1 gives an example of the program output which can help the accelerator physicist considerably to analyze the situation in case there are strange particles of mass  $M$  in the composition.

Obviously, <sup>207</sup>Pb<sup>9+</sup> and <sup>184</sup>W<sup>8+</sup> cannot be separated by standard analyzing systems. Therefore the use of tungsten has to be avoided as construction material in the ion source.

Table 1: Relevant numbers of the discussed example.

|    | $\zeta$ | $M$ | %     | $A$      | $A/\zeta$      |
|----|---------|-----|-------|----------|----------------|
| Ta | 8       | 181 | 99.99 | 180.9480 | 22.6185        |
| W  | 8       | 182 | 26.35 | 181.9483 | 22.7435        |
| W  | 8       | 183 | 14.32 | 182.9503 | 22.8688        |
| W  | 8       | 184 | 30.68 | 183.9510 | <b>22.9939</b> |
| W  | 8       | 186 | 28.49 | 185.9544 | 23.2443        |
| Pb | 9       | 204 | 1.54  | 203.9730 | 22.6637        |
| Pb | 9       | 206 | 22.62 | 205.9745 | 22.8861        |
| Pb | 9       | 207 | 22.62 | 206.9759 | <b>22.9973</b> |

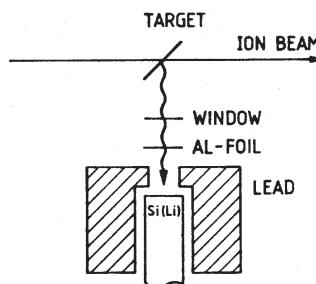


Figure 1: Schematic view of a set-up for X-ray spectroscopy taken from ref.[1].

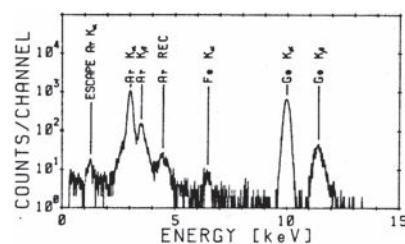


Figure 2: X-ray spectrum of a <sup>76</sup>Ge beam with mixtures of <sup>38</sup>Ar and <sup>57</sup>Fe taken from ref.[1].

As far as known from the experience there are the following similar cases:

- <sup>96</sup>Mo<sup>4+</sup> and <sup>144</sup>Sm<sup>6+</sup>
- <sup>40</sup>Ar and <sup>40</sup>Ca
- <sup>76</sup>Ge<sup>4+</sup>, <sup>38</sup>Ar<sup>2+</sup> and <sup>57</sup>Fe<sup>3+</sup>

To analyze the contents of the beam, a set-up based on x-ray spectroscopy may be used [1]. The ions are excited by a thin carbon foil and the emitted radiation is analyzed by a semiconductor

detector. A schematic view of the arrangement is shown in Fig.1 and Fig.2 shows a X-ray spectrum of  $^{76}\text{Ge}$  beam with mixtures of  $^{38}\text{Ar}$  and  $^{57}\text{Fe}$ , which could not be separated by the analyzing system of the injection area.

The interpretation of stripper spectra will be very difficult in such cases (see the example in section 7.4.4).

## References

- [1] U. Krause, PAC85 Trans. Nucl. Sci., NS 32, 5, 1941-1943, Oct. 85

### 7.4.2 Longitudinal Phase Space Measurement - Ion Beams P. Forck (GSI), P. Strehl (PET)

Considering rf-accelerators with particles moving in a bunch along the  $z$ -axis a single particle

- can have the correct kinetic energy, but may be a small time  $\Delta t$  ahead or behind the reference particle,
- can be in phase with the reference particle, but having a deviation in energy  $\Delta W$ ,
- or may even have deviations in  $\Delta t$  and  $\Delta W$ .

Therefore measurements of energy  $W$ , energy spread  $\Delta W/W$ , phase angle  $\Delta\phi$ , respectively  $\Delta t$  are very important for a correct setup of the rf-structures, including bunchers and de-bunchers.

Keeping in mind that  $dp_z/p_z \propto dW/W$  and  $dz = \beta c \Delta t = \beta c \Delta \phi / \omega$  the parameters are related to the longitudinal phase plane defined by  $A_z = \int \int dz dp_z$  as well as to the longitudinal emittance  $\epsilon_z = A_z/\pi = \frac{1}{\pi} \int \int d\phi dW$ .

Considering heavy ion machines, especially Linacs with  $\beta < 1$  measurement of momentum or energy can be performed nondestructively by a time of flight (TOF) measurement, which is based on a measurement of the two SI base units: length and time. Inserting two or three capacitive pick-ups as shown in Figure 1 and schematically in Figure 2 into the beam pipe, precise time of flight measurements can be performed, taking advantage of the periodic bunch structure. In the scheme with three pick-ups the time of flight  $t_c$  between  $P_2$  and  $P_3$  has to be measured, to determine the unknown number  $N$  of bunches between pick-ups  $P_1$  and  $P_2$ . To avoid ambiguity,

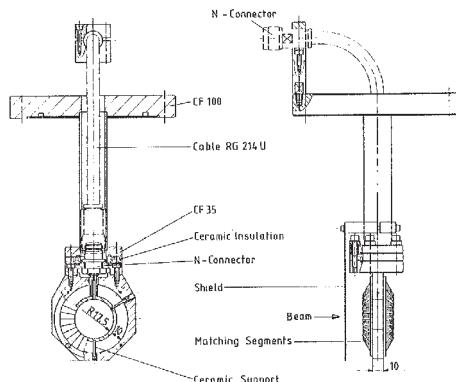


Figure 1: Capacitive pick-up in  $50\Omega$  geometry [1].

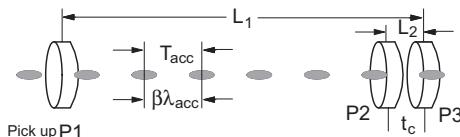


Figure 2: Setup for time of flight measurements [1].

the spacing between  $P_2$  and  $P_3$  has to fulfill the condition  $L_2 < \beta \lambda_{acc}$ . With the condition  $\Delta N < 1$  the required accuracy in the determination of  $t_c$  can be estimated from the relation  $N \Delta t_c / t_c < 1$ . The number of bunches  $N$  then is determined from:

$$\beta_c = \frac{1}{c} \frac{L_2}{t_c} \rightarrow N = INT\left(\frac{L_1}{\beta_c \lambda_{acc}}\right) \quad (1)$$

In the second step the measurement of the time difference between the arrival of a bunch at pick-up  $P_1$  and the first bunch in the train, arriving at  $P_2$  leads to  $\delta t$  and, finally to the time of flight  $t = NT_{acc} + \delta t$ . With  $\beta = L_1 / ct$  the kinetic beam energy  $W$  (and momentum  $p$ ) follows immediately from the well known relativistic relations between  $W, p, \beta, \gamma$ .

If the energy is known within certain limits (for example from the setting of an inflector magnet, spectrometer or just from the knowledge of the active accelerating structures), the coarse measurement can be left out, measuring only  $t_1$  so that only two pick-ups are needed. After specifying a lower and upper limit of possible energies it is recommended to calculate the relevant parameter as shown for example in table 1 to unequivocally select the correct  $\beta$ -value.

## Sec.7.4: BEAM INSTRUMENTATION AND DIAGNOSTICS

Table 1: Calculated beam energy in dependence of  $N$ . The data hold for an accelerating rf of 36.136 MHz, as typical for a heavy accelerator.

|   | $t$ [ns] | $\beta$ [%] | $\beta\lambda_{acc}$ [mm] | $W$ [MeV/u] |
|---|----------|-------------|---------------------------|-------------|
| 5 | 147.246  | 7.402       | 614.089                   | 2.562       |
| 6 | 174.919  | 6.231       | 516.937                   | 1.814       |
| 7 | 202.593  | 5.38        | 446.326                   | 1.351       |
| 8 | 230.266  | 4.733       | 392.686                   | 1.045       |
| 9 | 257.939  | 4.225       | 350.557                   | 0.833       |

In both cases the accuracy that can be achieved may be estimated nonrelativistically at:

$$\frac{\Delta\beta}{\beta} = \frac{\Delta v}{v} = \sqrt{\left[\frac{\Delta L}{L}\right]^2 + \left[\frac{\Delta(\delta t)}{NT_{acc} + \delta t}\right]^2} \quad (2)$$

where  $T$  is the period of the accelerating rf which in general is known with very high precision. Due to the sharp zero crossing of the pick-up signals it holds  $\Delta(\delta t) \lesssim 100$  ps. The capacitive pick-ups can also be used for a nondestructive measurement of the bunch length. Approximating the particle distribution in the  $\Delta W$ -,  $\Delta\phi$ - plane by ellipses, the longitudinal emittance characterized by the TWISS [2] parameters can be determined from nondestructive measurements by capacitive pick-ups. The simplest method is, to measure the bunch length at least at three pick-ups along a drift space. Taking the deviation in phase  $\Delta\phi$  and the energy spread  $\Delta W/W$  as the variables in the longitudinal phase plane, the coordinates after a drift  $L$  are determined from the matrix transformation:

$$\begin{pmatrix} \Delta\phi_1 \\ \Delta W_1/W_0 \end{pmatrix} = \begin{pmatrix} 1 & k \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \Delta\phi_0 \\ \Delta W_0/W_0 \end{pmatrix} \quad (3)$$

In non-relativistic approximation the coefficient  $k$  is easily derived from:

$$\Delta\phi = \omega\Delta t = -\frac{2\pi f L}{\Delta v} = -\frac{2\pi f L}{\Delta\beta c} \quad (4)$$

$$\frac{\Delta\beta}{\beta} = \frac{1}{2} \frac{\Delta W}{W} \rightarrow k [\text{rad}] = -\frac{\pi f L}{\beta c} \quad (5)$$

In practice the bunch length is measured on the time scale. Measuring the drift space in [m] and using [ns] as the unit on the time scale, the matrix equation 3 can be converted to

$$\begin{pmatrix} \Delta t_1 \\ \Delta W_1/W_0 \end{pmatrix} = \quad (6)$$

$$\begin{pmatrix} 1 & k^* \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \Delta t_0 \\ \Delta W_0/W_0 \end{pmatrix} \quad (7)$$

$$k^* [\text{ns}] = -\frac{1000 L [\text{m}]}{2\beta c [\text{mm/ns}]} \quad (7)$$

Therefore the bunch length  $\Delta t_1$  at position "1" is determined from the length  $\Delta t_0$  at position "0":

$$\Delta t_1 = \sqrt{\varepsilon\beta_1} = \sqrt{a_{11}^2 \varepsilon\beta_0 - 2 a_{11} a_{12} \varepsilon\alpha_0 + a_{12}^2 \varepsilon\gamma_0} \quad (8)$$

with  $a_{11} = 1$ ,  $a_{12} = k^*$ ,  $a_{21} = 0$  and  $a_{22} = 1$ . Thus, the Twiss parameters can be determined from three or more measured bunch lengths along a drift space by solving the corresponding equation system (see for example [1]). Considering the action of a rf-cavity is adjusted, like a single gap resonator or all kind of bunchers, on the particles leads to a further possibility to determine the TWISS parameters with even only one device for bunch length measurements. Adjusting the phase of the rf-cavity so that the reference particle arrives at the zero crossing of the rf ( $\phi = 0$ ) particles with small  $\Delta\phi$  around the reference particle gain energy given by  $\frac{\Delta W}{W} = \frac{1}{W} \frac{\zeta e}{A} T_{transit} U_0 \Delta\phi$ , with  $\zeta$  = charge state of the ions,  $A$  = their mass number,  $T_{transit}$  = transit time factor and  $U_0 = rf$ - voltage amplitude. Measuring again the bunch length  $\Delta t$ , the transfer matrix is given by:

$$\begin{pmatrix} \Delta t_1 \\ \Delta W_1/W_0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ DU_0 & 1 \end{pmatrix} \begin{pmatrix} \Delta t_0 \\ \Delta W_0/W_0 \end{pmatrix} \quad (9)$$

The constant  $D$  can be determined from the relation  $\frac{\Delta W}{W} = D U_0 \cos\phi_s = D U_0 \sin\phi$  by varying the phase of the rf-cavity, measuring the energy gain and calculating  $D$  with high precision from the slope in a diagram  $\Delta W/W = f(\sin\phi)$  as shown in Figure 3. Conversion to a phase plane with  $\Delta t$  as the variable of the abscissa has been performed replacing  $\sin\phi$  by  $\sin\phi/2\pi f$ . A practical dimension of  $D$  is [% / V ns]. After  $D$  has been determined the phase of the rf-cavity is to the zero crossing for the center of the bunch and the bunch length is measured in dependence of the rf-voltage. Taking more than three settings, the TWISS parameters can be calculated from a least squares fit as shown in Figure 4.

For non-relativistic beam energies, the electro-magnetic field caused by the bunched beam has a significant longitudinal component propagating faster than the bunch itself, see e.g. [3], [4]. With a pick-up this electro-magnetic field is sensed and the recorded signal does not necessarily represent the particle distribution within the bunch. An alternative method is based on the emission of secondary electrons from an

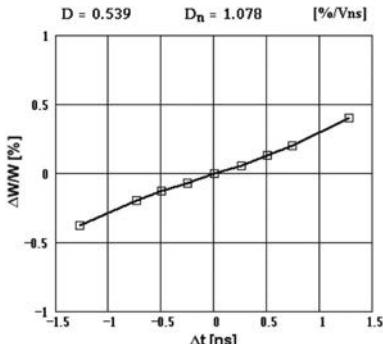


Figure 3: Determination of the characteristic cavity constant  $D$  from the slope  $\Delta W/W = f(\Delta t)$ .

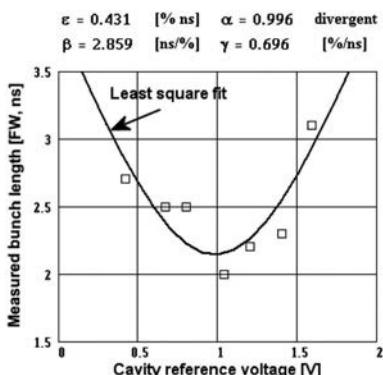


Figure 4: Determination of the longitudinal emittance from a least squares fit of the measured bunch length in dependence of the rf – voltage [1].

intersecting wire, as schematically depicted in Figure 5 and reviewed in [5]. The typical setup consists of a tungsten wire of 0.1 mm diameter intersecting the beam where secondary electrons are emitted within a time short compared to the bunch length. Due to the negative biasing of typically -10 kV these electrons are accelerated towards the pipe edge, where a small slit is positioned. The electrons pass a deflector plate driven by the bunch repetition frequency (phase axis), to convert the time information to different angles. After a drift of typically 30 cm the electrons location is monitored representing the bunch structure. A resolution of  $10^0$  with respect to the bunch repetition frequency can be achieved. To prevent the destruction of the wire by the large beam power at high current accelerators

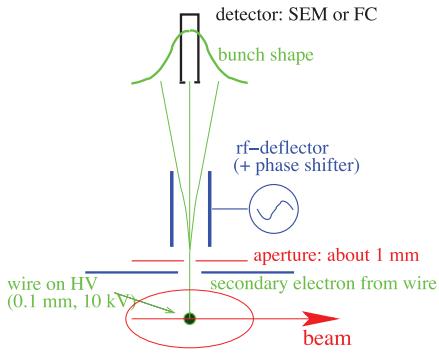


Figure 5: Scheme of a Bunch Shape Monitor based on the determination of the arrival time.

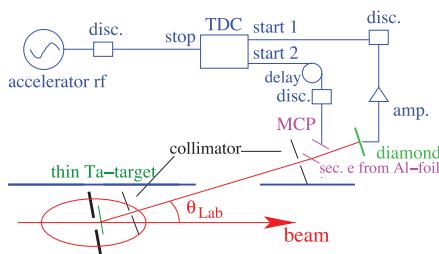


Figure 6: Scheme of a Bunch Shape Monitor based on arrival time measurement of individual particles.

the wire is inserted at the transverse edge of the bunches. This is a valid approach because at most accelerators the coupling between the longitudinal and transverse degree of freedom can be neglected.

A different approach is based on the determination of the beam particle's arrival time at a detector measured with respect to the bunch repetition time (corresponding to the phase axis). The best time resolution can be achieved if single particles are detected and an attenuation is required to have less than one particle per bunch. This can be performed by Rutherford scattering within a thin foil and a solid angle restriction by a collimator. In Figure 6 an example of such setup is shown [6] using a Tantalum foil and a scattering angle of  $\theta = 2.5^\circ$  for an attenuation of about  $10^{-4}$ . The resolution of  $\sim 30 \text{ ps}$  can be reached using fast single particle detectors like artificial diamond. Measuring the time-of-flight between two detectors for many individual beam ions, the longitudinal phase space distribution can be reconstructed.

## References

- [1] Strehl, P., Beam Instrumentation and Diagnostics, Springer-Verlag Berlin Heidelberg 2006, ISSN 1611-1052
- [2] Joho, W., Representation of Beam Ellipses for Transport Calculations, SIN-report, TM-11-14, (1980)
- [3] R.E. Schafer, BIW 1993, p.303
- [4] P. Kowina et al., DIPAC09, p.35
- [5] A. Feschenko, PAC01, p.517
- [6] P. Forck et al., DIPAC99, p.186 and P. Forck et al., LINAC2000, p.166

### 7.4.3 Heavy Ion Linacs - Emittance Measurements

*P. Forck (GSI), P. Strehl (PET)*

In the transversal phase space the two projected phase space areas are defined by  $A_y = \int \int dy dp_y$      $A_z = \int \int dz dp_z$ . In general for heavy ions in Linacs the motion in the transverse directions  $\dot{x}$  and  $\dot{y}$  are small so that relativistic effects can be neglected and it holds  $p_x = m\dot{x} = m_0\gamma \frac{dx}{dz} \frac{dz}{dt} = m_0\gamma\beta c \frac{dx}{dz} = m_0\gamma\beta cx'$  and in consequence  $A_x = m_0\gamma\beta c \int \int dx dx'$ . Because for most beams these areas can be well described by ellipses it is convenient to define  $\varepsilon_x = \frac{1}{\pi} \int \int dx dx'$  and  $\varepsilon_y = \frac{1}{\pi} \int \int dy dy'$  as the horizontal and vertical emittance. Therefore for a particle moving along the  $z$ -axis one has to determine the offset against  $x = 0$  ( $y = 0$ ) and the divergence angle  $dx/dz$  ( $dy/dz$ ). A simplified scheme based on a slit-detector system is shown in Figure 1. Typical parameters for the sandwich detector are:  $N = 32 - 64$  collector strips with  $d = 0.05 - 0.2$  mm and  $s = 0.2$  mm.

The unit may be mounted together onto a stepping motor driven feedthrough which results in a maximum measurable divergence  $x'_{max} = 1/2 N (d + s)/L$ . If slit and detector are mounted separately onto two feedthroughs, the achievable

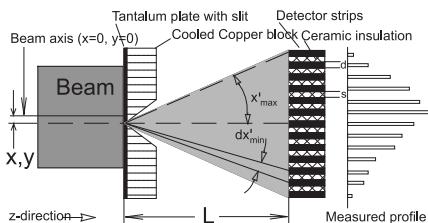


Figure 1: Scheme of a slit-detector system.

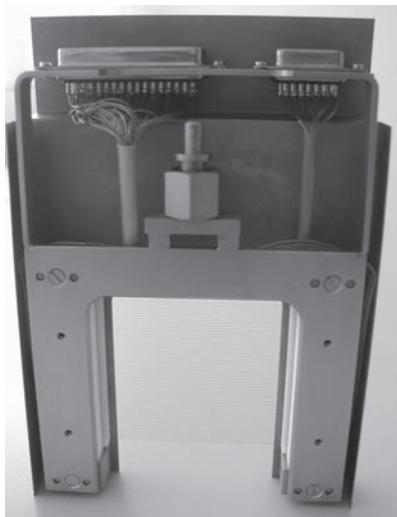


Figure 2: Harp detector: 60 W-Re wires, 0.1 mm in diameter, spacing 1 mm.

range  $x'_{max}$  can be increased by offset - positions of the detector, while the resolution in divergence can be improved by intermediate steps of the detector feedthrough. Instead of a detector sandwich as shown in Figure 1, a harp as shown in Figure 2 can be used.

**Signal Estimation** Referring to Figure 3, assuming a parabolic intensity distribution over the radial coordinate  $i(R_s) = (2I_0/\pi R_b^2)(1 - (S_x^2 + y_s^2)/R_b^2)$ , in a good approximation the current  $I_s$  passing the slit at position  $S_x$  is

$$I_s(S_x) \approx \quad (1)$$

$$2\delta S \frac{2I_0}{\pi R_b^2} \int_0^{\sqrt{R_b^2 - S_x^2}} \left(1 - \frac{S_x^2 + y_s^2}{R_b^2}\right) dy_s$$

$$I_s(S_x) \approx \frac{8\delta S I_0}{3R_b \pi} \left(1 - \frac{S_x^2}{R_b^2}\right)^{3/2}. \quad (2)$$

Defining  $V = S_x/R_b$  with  $0 \leq V \leq 1$ , the percentage passing the slit is approximately  $P(V) \approx 100(8\delta S/3\pi R_b) (1 - V^2)^{3/2}$ . With this definition  $V = 0$  corresponds to the center position of the slit, while for  $V = 1$  the slit is at the boundary of the beam. Figure 4 gives the calculated percentage for a slit width of 0.1 mm. The resulting percentage hitting a collector strip depends on the width of the strips, the divergence in the beam and the distance between slit and detector (for an estimation see e.g. [1]).

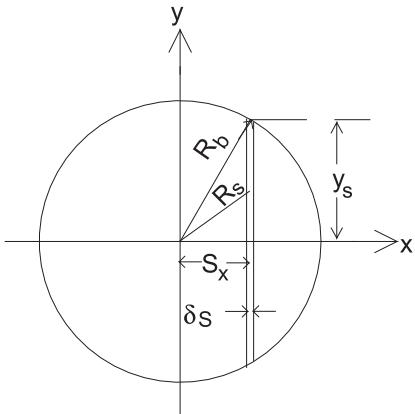
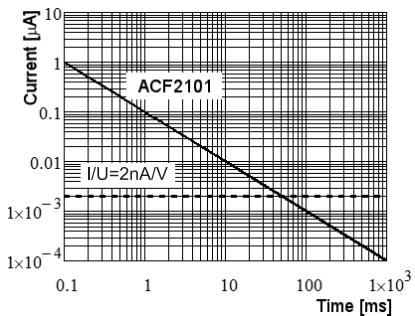


Figure 3: Scheme to estimate intensity behind a slit.

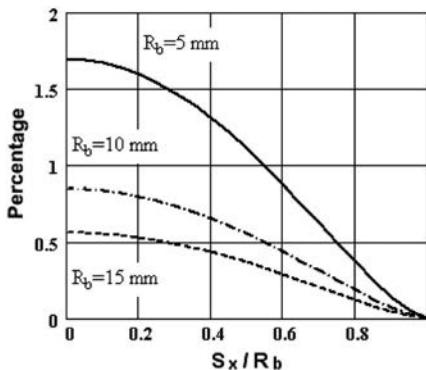
Figure 4: Percentage of beam passing a  $0.1\text{ mm}$  slit.

**Signal processing** Conversion of collector currents  $I$  into reasonable voltages  $U$  for digitization can be performed either using an operational amplifier as  $I/U$ - converter with  $U(t) = -I(t)R$ , or a switched integrator with  $U(T) = -\frac{1}{C} \int_0^T I(t)dt$ . Figure 5 the most sensitive range of an  $I/U$ -converter with  $I_U = I/U = 2\text{ nA/V}$  with an IC ACF2101 having a conversion rate of  $Q_U = Q/U = 10^{-10}\text{ As/V}$ .

**Evaluation of Data** Each measured point corresponds to a phase space area  $\Delta x \Delta x' (\Delta y \Delta y')$ , determined by the step width of the slit and the width of the detector strip. So, the 100%- emittance  $\epsilon_x$  is given by

$$\epsilon_x = n_x n_{x'} \Delta x \Delta x' / \pi, \quad (3)$$

with  $n_x$  as the number of slit positions and  $n_{x'}$  as the number of measured values of  $x'$ . It's usual to calculate the emittance in dependence of a given

Figure 5: Comparison of an  $I/U$ - converter with a switched integrator.

percentage. From the evaluation software the measured intensities for the area elements  $\Delta x \Delta x'$  are given in a matrix  $I_{n_x, n_{x'}}$  having  $n_x$  rows and  $n_{x'}$  columns. Different algorithms are applied to  $I_{n_x, n_{x'}}$  to derive an emittance representing a given percentage of a beam:

**Emittance mode** Summing up all intensities  $I_{n_x, n_{x'}}$  defines the maximum intensity  $I_{max}$ , taken as reference to calculate the percentages. Therefore, to calculate for example the so-called 80%- emittance a value of  $S = 0.2 I_{max} / n_x n_{x'}$  has to be subtracted from each element with  $I_{n_x, n_{x'}} > 0$ . No subtraction takes places if  $I_{n_x, n_{x'}} = 0$ . Furthermore since  $I_{n_x, n_{x'}} < 0$  is not allowed, the subtraction has to be performed successively in much smaller steps as defined by  $S$  (for example  $S/100$ ). Elements coming below zero are set to zero. The procedure is repeated until the required value of  $S n_x n_{x'}$  is achieved within a given threshold.

**Intensity mode** The maximum intensity of all  $I_{n_x, n_{x'}}$  gives the reference for the percentages. The value calculated for a chosen percentage is subtracted from each element. Elements becoming below zero are set to zero.

**Minimum subtraction mode** The reference value for the percentage is again the sum over all intensities. But, to subtract the calculated amount from each element the smallest values are removed (set to zero) as long as the required value is achieved. No subtraction takes places from elements with large  $I_{n_x, n_{x'}}$ - values. If it happens,

that the last subtraction step is larger as the required one, the resulting difference is added to the last processed element.

The discussed slit - detector system and similar systems like a multi slit system or pepper pot allow the determination of intensity distributions of arbitrarily shape in the transverse phase planes. This will be necessary to study aberrations as well as higher order effects of beam transport elements, to investigate and improve ion sources, to study space charge effects and during commissioning of a new machine. On the other hand, approximating the measured emittance patterns by ellipses, characterized by the TWISS parameters are much more easy to handle with respect to: (i) the transformation of a beam through a channel of various beam transport elements, (ii) the comparison of emittances and acceptances, including correct matching of accelerator sections, (iii) the elaboration and store of data sets for computer aided routine operation, including optimization procedures.

#### Determination of TWISS Parameters

Determination of rms-values of a two-dimensional Gaussian normal distribution is one of the most applied methods. The relevant formulas are:

$$Sum = \sum_{n_x} \sum_{n_{x'}} I_{n_x, n_{x'}} \quad (4)$$

$$X_0 = \frac{1}{Sum} \sum_{n_x} \sum_{n_{x'}} n_x I_{n_x, n_{x'}} \quad (5)$$

$$X'_0 = \frac{1}{Sum} \sum_{n_x} \sum_{n_{x'}} n_{x'} I_{n_x, n_{x'}} \quad (6)$$

This gives the numbers for the center in the matrix. With  $\Delta x = (x_{\max} - x_{\min})/n_x$  and  $\Delta x' = (x'_{\max} - x'_{\min})/n_{x'}$  the  $\sigma$ - values of the Gaussian distribution are given by:

$$D_x = \frac{\Delta x^2}{Sum} \sum_{n_x} \sum_{n_{x'}} [n_x - X_0]^2 I_{n_x, n_{x'}} \quad (7)$$

$$D_{x'} = \frac{\Delta x'^2}{Sum} \sum_{n_x} \sum_{n_{x'}} [n_{x'} - X'_0]^2 I_{n_x, n_{x'}} \quad (8)$$

$$M_{xx'} = \frac{\Delta x \Delta x'}{Sum} \sum_{n_x} \sum_{n_{x'}} [n_x - X_0][n_{x'} - X'_0] I_{n_x, n_{x'}} \quad (9)$$

$$\sigma_x = \sqrt{D_x}, \quad \sigma_{x'} = \sqrt{D_{x'}} \quad (10)$$

$$\rho = \frac{M_{xx'}}{\sigma_x \sigma_{x'}} \quad (10)$$

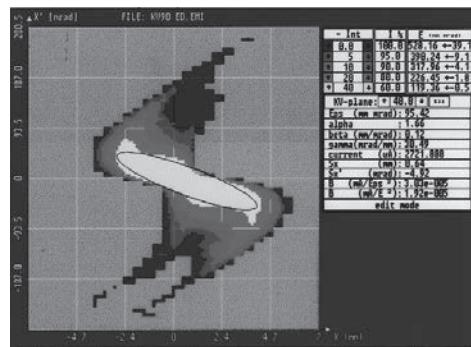


Figure 6 shows an extremely deformed emittance area, but even in this case the fitted ellipse contains about 60% of the beam intensity.

**7.4.4 Charge State - Ion Beams**  
P. Forck (GSI), P. Strehl (PET)

Determination, separation or controlled change of the charge state  $\varsigma$  e.g. by stripping may be necessary for the following reasons: to evaluate and optimize ion sources, for analyzing stripper spectra, for determination of particle momentum [see also 7.4.2], for charge-change extraction [for example in cyclotrons, storage rings], for minimizing charge exchange in e-coolers, for increasing the output energy of tandem accelerators, for the evaluation of atomic physics experiments.

Considering heavy ion accelerators, typical ion sources, such as sputtering ion sources of the PIG-type or ECR's will produce a spectrum of charge states for each isotope as shown for example in Figures 1 and 2.

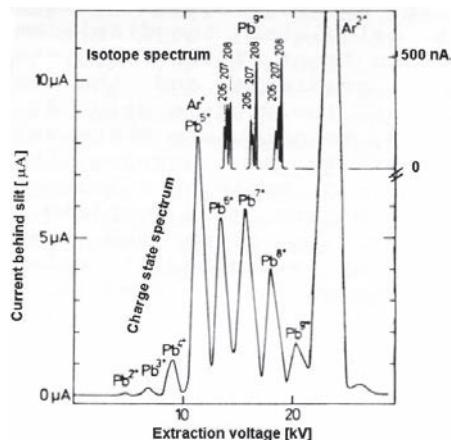


Figure 1: Charge state and mass spectrum (inset) of lead from a PIG, using Ar as auxiliary gas.

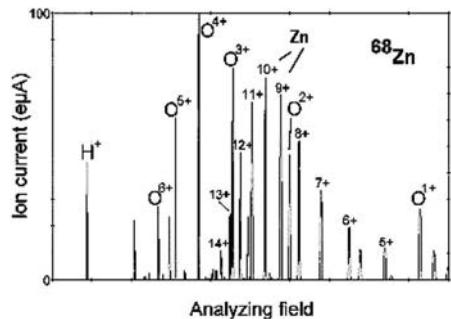


Figure 2: Spectrum from an ECR.

In general, the separation is based on the relation of the magnetic rigidity and electrostatic rigidity. Mostly magnetic separation is used.

$$B\rho = 3.10715 \frac{A}{\zeta} \beta \gamma \quad (1)$$

magnetic rigidity

$$Ep = \beta c B\rho \quad (2)$$

electrostatic rigidity

with  $A$  as the mass number and  $\rho$  as the bending radius. Therefore

$$B_1/B_2 = \zeta_2/\zeta_1 \quad \text{if } \beta \text{ is constant} \quad (3)$$

$$B_1/B_2 = \sqrt{\zeta_2/\zeta_1} \quad \text{if } T \text{ is constant} \quad (4)$$

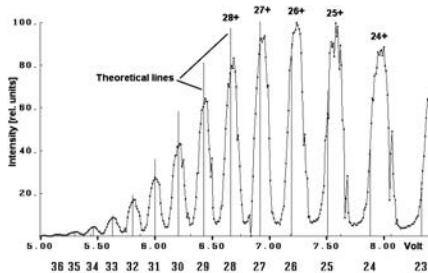


Figure 3: Charge state spectrum (Gas) of Uranium, projectile energy 1.4 MeV/u.

where  $T = \zeta eU$  is the kinetic energy,  $U$  = accelerating voltage and  $\beta \ll 1$ . The geometric separation of lines with different charge states by a magnetic analyzing system depends generally on the dispersion  $D = (\Delta y/R)/(\Delta p/p) = (y|\delta|)$ , on the magnification  $M = (y|y|)$  and of course on the widths of the analyzing slits. In [2] the resolving power  $RP$  is defined as:

$$RP \equiv \frac{p}{\Delta p_{1/2}} = \frac{RD}{MS}, \quad (5)$$

here  $S$  is the width of the entrance slit, which cannot be made arbitrarily small due to the decreasing intensity. Similar relations hold also for electrostatic and more sophisticated analyzing systems. Separation of the Pb-isotopes (see Figure 1) requires a resolution  $M/\Delta M \sim 200$ .

To reduce the required rf-power, gas or foil strippers are used to increase the charge state  $\zeta$ . Determination of the unknown charge states of a stripper spectrum as shown from a gas stripper in Figure 3 and a foil stripper in Figure 4 may be performed by applying least squares fits to the measured data. Assuming magnetic separation according to relation (1) the following fits may be applied: If the energy of the particles is known

$$\sum_i \left[ B_i - \frac{3.10715 A \beta \gamma}{\rho \zeta_i} \right]^2 = Min \quad (6)$$

and if energy is not known:

$$\sum_i \left[ \frac{B_i}{B_k} - \frac{\zeta_k}{\zeta_i} \right]^2 = Min \quad (7)$$

or

$$\sum_{i,k} \left[ \frac{B_i - B_{i+k}}{B_k + B_{i+k}} - \frac{\zeta_{i+k} - \zeta_i}{\zeta_{i+k} + \zeta_i} \right]^2 = Min \quad (8)$$

In Figures 3 and 4 the assignments of the charge states have been determined in this way.

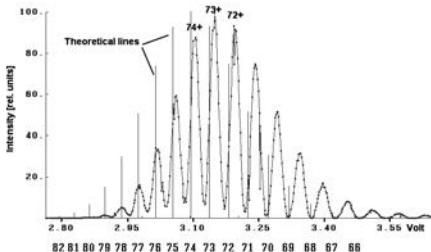


Figure 4: Charge state spectrum (Foil) of Uranium, projectile energy 11.5 MeV/u.

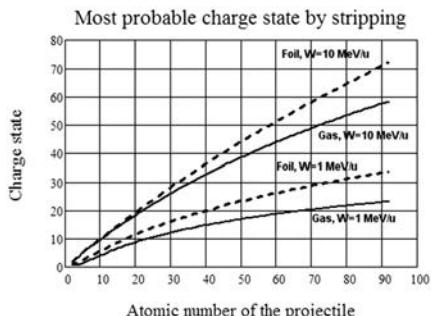


Figure 5: Calculation of the most probable charge state by stripping according to equation 10.

More generally, if charge states and energy have to be determined from the measured spectra the least squares fit

$$\sum_i \left[ B_i - \left( \frac{a}{\zeta_i} + b \right) \right]^2 = \text{Min} \quad (9)$$

can be applied. Here the fit-variable  $a$  leads to the energy via the relation  $\beta\gamma = ap/(3.10715A)$ , while the fit-variable  $b$  describes a magnetic field offset which may result from incorrect injection into the magnet. Employing the semi-empirical formula [1]

$$\bar{\zeta} = Z \left[ 1 - Ce^{-137\beta Z^{-\gamma}} \right] \quad (10)$$

the most probable charge states  $\bar{\zeta}$  behind a stripper target (gas or foil) can be calculated, assuming that the target thickness  $x$  corresponds to the equilibrium charge state distribution ( $dN_\zeta/dx = 0$ ). For the energy range  $1 \leq W \leq 20 \text{ MeV/u}$  the constants  $C$  and  $\gamma$  have been determined experimentally at the UNILAC (GSI) to :  $C = C_{CB} + 140/Z^2$ , with  $C_{CB} = 1.0285$ , and  $\gamma = 0.56$  (foil),  $\gamma = 0.65$  (gas). Figure 5 gives calculated examples for gas and foil strippers.

## References

- [1] Heckmann, H. H., Hubbard, E. L., Simon, W. G., PR, 129, 1240, (1963)
- [2] Septier, A., Focusing of charged particles, Volume II, Academic Press Inc., London, 1967

## 7.4.5 Beam Current Measurement

*J. Hinkson, LBNL*

**Nonintercepting methods** [1] Here we describe commonly used magnetic devices.

**DCCT:** The dc current transformer, also known as zero flux current transformer, 2nd harmonic magnetic modulator and parametric current transformer, measures the dc component of the beam current. Fig.1 is a simplified schematic.

In it a high permeability toroidal magnetic core surrounds the beam, coupling to its magnetic field. In the absence of a beam a magnetic modulator periodically drives the core into positive and negative saturation, and the sense winding produces a perfectly symmetrical positive and negative output voltage having no even harmonics. Magnetic field from the beam, or any other source that couples to the core, causes the core  $B$ - $H$  loop to become slightly offset, resulting in generation of a 2nd harmonic in the sense winding. This signal is filtered, rectified, amplified and feedback to a third winding which just cancels the disturbing flux from the beam. A precision resistor in series with this bucking winding then produces a voltage proportional to beam current. Being a null measurement the linearity and accuracy of the method are high.

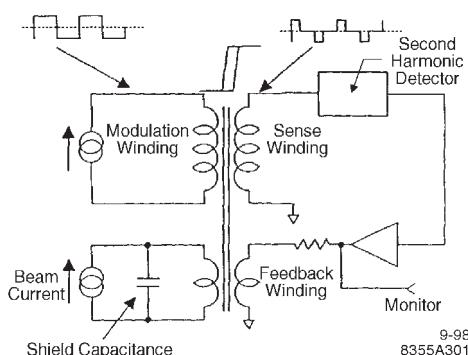


Figure 1: Typical DCCT arrangement.

To avoid excitation of the core by extraneous fields, it is customary to insert an insulating (ceramic) break in the beam pipe around which the core is placed. An electrical shield, well connected to the beam pipe on either side of the core, forces extraneous wall currents safely around the core so that only beam current links the core. Magnetic shielding around the whole insures that extraneous magnetic fields are excluded from the core as well. A carefully shielded and terminated single wire paralleling the beam through the core and driven by a precision current source calibrates the system.

Frequency response of this circuit is limited to half the modulating frequency, typically limited to 10 kHz by the magnetic material. By adding auxiliary, unmodulated, cores and another high gain amplifier combined with the DCCT, it is possible to extend the frequency response to hundreds of MHz [2]. Note that short beam bunches have components extending to 10 GHz which can excite resonances in the DCCT components and cause spurious outputs or cause heating damage, particularly to the magnetic material of the core. Shunting the break with a low inductance capacitor can help. No resistor less than  $10\text{ k}\Omega$  should be put across the gap. Conductivity of cooling water can be a problem. Radiation damage is also a concern. Cores having radiation hardened components are commercially available [3]. Complete DCCT units of excellent performance can be obtained commercially [3, 4]. Typical performance parameters are shown in Tab.1.

| Table 1 DCCT performance parameters |  |
|-------------------------------------|--|
| Freq. response                      | dc to 100 kHz                              |
| Full scale range                    | $\pm 1\text{ mA}$ to $\pm 10\text{ A}$     |
| Resolution                          | $0.5\text{ }\mu\text{A}$ (1 s integration) |
| Dynamic range                       | $>2 \times 10^7$                           |
| Absolute accuracy                   | better than 0.05%                          |
| Linearity                           | better than 0.01%                          |
| Core inner diameter                 | 50 to 250 mm                               |

**Current transformers:** For measuring pulsed currents, transformers are in wide use and a variety are commercially available [3, 6, 7]. Sensitivities up to 5 V/A are common. In selecting the V/A and Amp seconds of a unit, one must take care not to exceed the voltage rating of the windings or to saturate the core. Measurement of currents  $<100\text{ }\mu\text{A}$  by this method is made difficult by thermally and acoustically generated noise. Rise-times  $\sim 1\text{ ns}$  are very difficult to achieve because

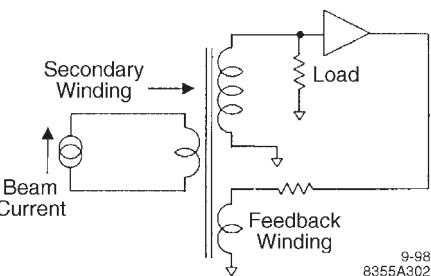


Figure 2: Current transformer with Hereward feed-back.

of stray capacitances from winding to winding and to ground. Nevertheless, some commercially available transformers have rise times less than 1 ns [3]. Eddy current losses in the core increase rapidly with frequency requiring use of very thin magnetic material (e.g.  $10\text{ }\mu\text{m}$  [5]).

While an ideal transformer should be insensitive to beam position and bunch length, at high frequencies both effects are present. The windings on an unshielded current transformer can couple to the electric field of the beam which may contaminate the measurement. Thus it is important to install an electrostatic shield surrounding the winding with insulating layer so as not to form a shorted turn around the core.

The low frequency response of the current transformer is determined by its droop rate  $1/\tau = R/L$  where  $R$  is the secondary circuit resistance and  $L$  its inductance. The low frequency response can be extended by use of the Hereward feedback technique shown in Fig.2. A drawback to this technique is increased noise.

**Integrating current transformer:** The ICT is a special application of the current transformer and is used to measure charge of bunches as short as 1 ps and as long as 1  $\mu\text{s}$ . Fig.3 shows a schematic.

The charge induced by short beam bunches is initially stored in a capacitor which then discharges into the transformer primary, spreading the width and reducing the amplitude. While temporal information about the bunch is lost, the charge is faithfully measured. The ICT output can be integrated as shown or displayed on an oscilloscope with integrating function.

**Wall current monitor:** Fig.4 shows a WCM as installed in the ALS [12].

An insulating gap forces the image currents in the wall to pass a belt of low inductance, low

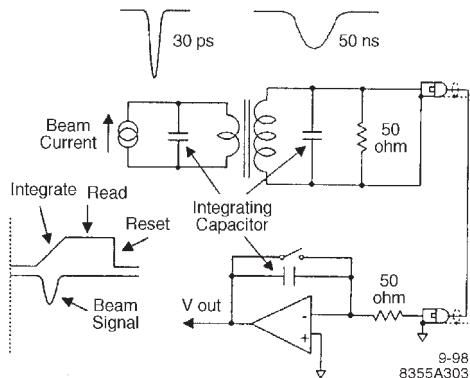


Figure 3: ICT and gated integrator.

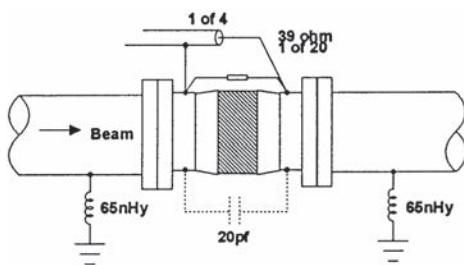


Figure 4: A simple wall current monitor.

value resistors arranged circumferentially about the gap. The output signal can be sensitive to beam position if the bunch length is comparable to WCM dimensions. In fact this detector can be used as a position monitor for times short compared to time for charge to equalize around the conducting beam pipe ends. If wanted this effect can be minimized by tapping the signal at several places around the perimeter and combining the signals in a broad band hybrid. The upper frequency limit (-3 dB) is  $\omega_{\text{high}} = 1/RC$  where  $R$  is the parallel combination of the belt of resistors and  $C$  is the gap capacitance. 2 GHz response is easy to achieve. A WCM of 6 GHz response has been demonstrated [8]. The low limit is  $\omega_{\text{low}} = R/L$  where  $L$  is the inductance shunting  $R$ . The low frequency response can be improved by placing ferrite cores on the beam pipe between the grounds and the resistor belt.

Wakefield contamination of the WCM monitor can be reduced by placing microwave absorbing material up and down stream of the WCM [8].

### Intercepting instruments

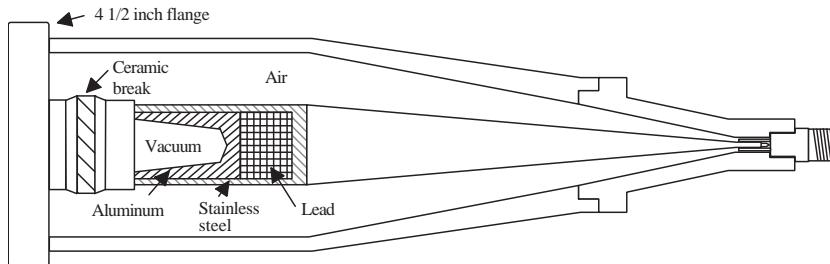
**Faraday cup:** Accelerated particles are stopped inside the cup and accumulated charge is converted into a corresponding current. Because all beam energy is deposited in the cup, its physical appearance will vary with the beam energy and power involved. Here we discuss cups suitable for up to a few hundred MeV and a few kW. Such devices are commercially available [9, 10, 11]. Some of them are equipped with actuators for removing them from the beam path. Important design factors are beam energy, beam power and power density, high frequency response and secondary electron emission. Target thickness must exceed range of the particles. High  $Z$  material such as Ta is often used. From 10 - 100 W, forced air cooling may suffice. Above that, cooling with low conductivity water is required. One commercial supplier [10] recommends that power density not exceed 140 W/mm<sup>2</sup> on a 1 kW unit. The internal geometry should distribute the power over as large an area as practical. Good high frequency performance entails good impedance matching. 50  $\Omega$  construction is common. An example from the ALS [12], shown in Fig.5, is used to measure a 50 MeV, 50 W beam.

A notable feature of this design is the commercial vacuum break supporting the beam target and coaxial center conductor. The coax is air insulated so that the cable connector need not be vacuum rated. A disadvantage is the relatively large capacitance introduced by the ceramic gap limiting the frequency response to 1 GHz. Industry offers cups with up to 2 GHz response.

Secondary emission must be suppressed for accurate measurements. This is done by making the target reentrant and introducing a bias of several hundred volts relative to the center conductor and/or introducing a dipole field of a few hundred gauss in the target region.

### References

- [1] R.C. Webber, Charged Particle Beam Current, AIP Proc. 333 (1994)
- [2] K.B. Unser, IEEE Trans. Nucl. Sci., NS-16 (1969)
- [3] Bergoz Precision Beam Instrumentation, 01170 Crozat, France
- [4] Holec Projects B.V., Hengelo, Netherlands
- [5] K.B. Unser, PAC89
- [6] Pearson Electronics Inc., Palo Alto CA.

Figure 5:  $50 \Omega$  Faraday cup.

- [7] Ion Physics Corp., Atkinson NH
- [8] R.C. Webber, Longitudinal Emittance, AIP Proc. 212 (1989)
- [9] Kaliper Co., Lee Hall VA
- [10] National Electrostatics Co., Middleton WI
- [11] Princeton Scientific Corp., Princeton, NJ
- [12] J. Hinkson, Advanced Light Source Beam Diagnostics, AIP Proc. 319 (1993)

#### 7.4.6 Beam Position Monitors

*J.A. Hinkson, LBNL  
S. Smith, SLAC & CERN  
G. Decker, ANL*

Beam position monitors nominally measure transverse location of the center of beam charge, often reporting total charge as well. There is great diversity in monitoring techniques appropriate to the wide variety of beam characteristics. Requirements vary dramatically between machines: light sources demand orbit stability while linacs often stress bandwidth, *e.g.* single bunch capability. Particle physics accelerators have tight requirements on stability over wide ranges of currents and fill patterns. A BPM system consists of pickups and a processor (receiver).

**Nonintercepting methods** There typically are several nonintercepting BPMs per betatron wavelength. Wideband BPMs and receivers measure position of individual bunches while narrow-band receivers are used for precision closed orbit measurement and feedback control. Conflicting requirements for these two functions may require the use of separate systems. Optimum measurement bandwidth depends on the signal spectrum; for position measurements averaging over closely-spaced bunches signal-to-noise ratio improves at narrow bandwidths, whereas for an isolated single bunch signal-to-noise improves at high bandwidth.

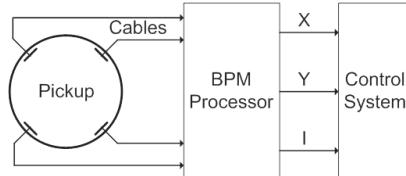


Figure 1: Typical BPM pickup and processor.

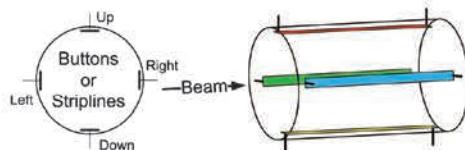


Figure 2: Typical Buttons and striplines.

**Pickups** Common beam pickup sensors [1] are buttons, striplines, and resonant cavities. Rings usually use button pickups for low beam impedance. Buttons are mounted flush with or slightly recessed from the beam duct wall. At low processing frequencies ( $f \ll c/d$  for button diameter  $d$ ) coupling is proportional to button area, quite low for bunch lengths greater than the button diameter. Figure 1 shows buttons rotated to avoid direct synchrotron radiation.

Linacs and transfer lines often employ striplines for higher signal-to-noise at the cost of higher beam impedance. [2] The end of the stripline away from the feedthrough may be shorted, open, or terminated in the characteristic impedance. For matched termination the pickup is directional; for relativistic beam no signal exits the downstream feedthrough. This directionality may be exploited to separate the signals of counter-propagating beams.

Resonant cavity position monitors are used for highest resolution in linacs. [3] Output couplers extracting only the dipole (position-sensitive) cavity modes avoid the complexity and dynamic range issues of computing position from small differences of large beam signals. Cavity BPMs are particularly suited for linac-based FELs requiring sub-micron single-shot resolution and stability to maintain the strict colinearity needed for lasing. This was demonstrated at LCLS (SLAC) using copper X band cavity BPMs developed at Argonne National Laboratory. [4] A steel cavity operating at C band has been developed at the SPring-8 XFEL [5] and is the basis for the cavity BPMs for the European XFEL undulator section. [6] Cavity position monitors can potentially measure beam angles and bunch tilts ( $x - z$  or  $y - z$  correlations within the bunch). [7]

Other position sensitive detectors include split cylinders, magnetic loops and resistor belts arranged around a ceramic break. The latter are useful in induction accelerators with very high peak current. Detectors which image synchrotron light, transition, diffraction radiation, or X-rays are also used to measure beam position.

**Button Transfer impedance** For beam current  $I_b$  at frequency  $\omega$  and induced pickup voltage  $V_p$ , the transfer impedance is  $Z_t(\omega) = V_p/I_b$ . The transfer impedance of a button of radius  $a$  to a centered beam in a round beam pipe of radius  $r$  is

$$Z_t(\omega) = \frac{a^2\omega}{2r\beta c} Z(\omega) \text{ where } Z(\omega) = \frac{R/i\omega C}{R + 1/i\omega C}$$

$R$  is the load impedance and  $C$  the button capacitance, usually a few pF. The 3 dB corner frequency is  $\omega = 1/RC$ . Caution: narrow-band beam impedance occurs at wavelengths near the button circumference. [8]

**Stripline Response** Designing a stripline with desired transmission impedance may be done via 2-D electrostatic calculations or by approximation of the stripline as a planar microstrip. [9] The impulse response is a doublet of zero integral. For a stripline of length,  $L$ , shorted at the far end, the response to beam current  $I_b(t)$  is [10]

$$V(t) = \frac{\varphi Z}{4\pi} \left[ I_b(t) - I_b \left( t - \frac{2L}{c} \right) \right] \text{ where } \phi$$

is the azimuthal extent of the strip,  $Z$  is the stripline impedance. This response has periodic nulls at frequencies which are multiples of  $F_0 = \frac{c}{2L}$  and has peaks between. It may be flattened by tapering the strip, temporally spreading the echo

pulse and smoothing the frequency response. The spacing  $d$  must be tapered as well for constant impedance giving practical limits to this technique of about two octaves. [11] One can do even better at the price of sensitivity by tapered slot coupling to the strip or other transmission line exterior to the beam pipe. [12]

**Beam signals** The receiver bandwidth needed depends on the bunch spectrum. For a Gaussian bunch where  $\sigma$  is the rms bunch length the current at the  $n^{\text{th}}$  harmonic of the bunch frequency is given by:

$$I_b(t) = \frac{Q}{\sqrt{2\pi}\sigma} e^{\frac{-t^2}{2\sigma^2}}$$

$$I_n(\omega) = I_{avg} \cdot \begin{cases} 1 & n = 0 \\ 2e^{-(n\omega\sigma)^2/2} & n > 0 \end{cases}$$

Multiplying by the coupling impedance of the button or line gives the voltage at this harmonic.

**Position** Beam position is usually determined by the  $\Delta/\Sigma$  or “difference over sum” method. In the simple case of a round beam duct having small, upright pickups (buttons or striplines on the horizontal and vertical axes), a first-order estimate of beam position is given by [13]

$$X = \frac{R V_{left} - V_{right}}{2 V_{left} + V_{right}} \quad Y = \frac{R V_{up} - V_{down}}{2 V_{up} + V_{down}}$$

For other duct geometries this can be generalized by rotations and scaling with coefficients calculated analytically, for example by conformal mapping, by 2-D electrostatic solver, or by calibration with antennae or wires. [14] The above position formulae provide a good first-order approximation for small beam displacements; terms higher order in  $\Delta/\Sigma$  are sometimes included for improved linearity at large beam displacements.

**Pickup hardware** Buttons and feedthroughs with broad band response are available from several manufacturers. [15] Striplines are usually manufactured in-house with commercial 50  $\Omega$  feedthroughs.

**Operating frequency, narrow-band systems** Narrow-band systems are tuned to a selected revolution or bunch-spacing harmonic. Factors involved in frequency choice are transfer impedances of the pickups, cable losses, rf interference and component availability. The lowest practical frequency in storage ring applications is the rf frequency in the case that all buckets are full. Choosing a harmonic of the rf system frequency is usually satisfactory since transverse position information is contained in all sidebands

about its harmonics. If the harmonic is too high, the signal amplitude will be affected by bunch length changes. The highest frequency is set by the waveguide cutoff of the beam pipe. Care must be taken that useful features of the beam pipe, such as synchrotron radiation antechambers, do not lower the waveguide cutoff frequency below the BPM processing frequency as non-local beam signals can then confound BPM measurement.

**Calibration and alignment** A number of methods have been devised for “absolute calibration” referencing the BPM electrical center to fiducials on quadrupoles via precision mechanical movers exploring the beam aperture with antennae or coaxial wires. While these are useful in getting the location of electrical centers, beam based alignment schemes (Sec. 4.3.1, 4.7.2, 4.7.5) are far superior and permit frequent recalibration with great precision and accuracy.

**Signal processing methods** A number of factors determine the method to be chosen, single or combined function pickups, cable costs, electronics costs, bandwidth needed, resolution and accuracy, and required dynamic range among others. See the Beam Instrumentation Workshop, DIPAC, EPAC, and PAC proceedings in the references. One may categorize the many approaches in three broad classes.

Difference over sum The most commonly used method, can be effected in broad band, bunch by bunch versions and in narrow-band tuned receiver versions. Electronics may be multiplexed to avoid gain matching errors. In narrow-band systems great use is made of widely available rf and microwave IC’s. In such a system a band pass filter permits the passage of the chosen beam harmonic which is then mixed to form an IF signal for amplification and digitization. Multiplexed  $\Delta/\Sigma$  receivers can eliminate the differential errors inherent in the use of individual receivers. Normalization to beam intensity may be done using the signal in an automatic gain control loop. Intensity dependence in multiplexed  $\Delta/\Sigma$  receivers is very low, making them ideal position detectors for closed orbit feedback systems. Resolution and stability on the few  $\mu\text{m}$  level are achievable and dynamic ranges of 40 dB are not uncommon. Analog sum and difference signals can be generated in passive hybrids but great care must be taken in phase and amplitude matching at hybrid inputs.

AM to PM Amplitude differences in beam pickup signals can be translated into phase differences [16] or time differences. [17] These phase modulated signals are stripped of their amplitude components in limiters and compared in a phase detector whose output is proportional to beam displacement. Normalization to beam intensity is automatic. Bandwidth in these receivers is as much as 10% of the beam harmonic frequency making them useful for single turn measurements. Most often used with single function pickups, AM to PM detectors may be used with combined function pickups if  $180^\circ$  hybrids are used in front to form the  $\Delta$  and  $\Sigma$  signals but great care is needed in matching filters and cables. A good commercial hybrid with octave bandwidth may have 40 dB isolation between  $\Delta$  and  $\Sigma$  ports. This means that for centered beam the  $\Delta$  port signal may be 1% of the  $\Sigma$  port signal, not zero as we should wish for a good BPM.

Log ratio detectors Inexpensive logarithmic amplifier integrated circuits have led to the development of the log ratio detector. The principal advantages are normalized real time response, excellent dynamic range, up to 95 dB, for intensity and position measurements and circuit simplicity. Phase matched cables are not required. When used with pickups in a circular beam pipe the log ratio detector provides superior linearity by compensating to some degree the inherent non-linearity of the pickups. Bandwidth can be quite high, up to 100 MHz, making it useful for brief bursts of beam.

Digital BPM Processor Fast digitizers with sufficient resolution to acquire BPM signals are popular in new designs. Currently 16-bit ADCs are available at sample rates up to 125M/sec, either as ICs or packaged in digitizer modules. BPM button or stripline signals may be bandpass-filtered then digitized, either with or without analog downconversion before the ADC.

Ringing waveforms are digitized and the amplitude of each reconstructed from the time-series of samples. This digital processing includes

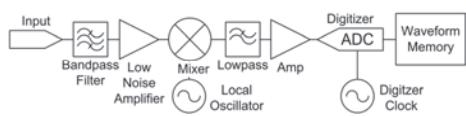


Figure 3: A channel of “digital BPM” including optional downconversion.

downconversion and filtering. Digital downconversion is done by multiplying the sequence of ADC samples by a complex digital “local oscillator” then lowpass-filtering, keeping the baseband signal while removing the sum frequency term. Position is then calculated from amplitudes by a conventional difference over sum algorithm. Performing most of the processing digitally provides several advantages, including [18, 19, 20]:

- Matched filtering, done digitally with identical coefficients on all channels. This reduces position offsets and their variation.
- Relaxed timing as the numerical algorithm can be insensitive to beam arrival time.
- Reduction of digitizer noise by averaging over many samples of the waveform, referred to as “processing gain”.
- Simultaneous acquisition at multiple frequencies or bandwidths. Allowing, for example, continuous calibration with a tone offset in frequency along with the beam signal.
- Software-selectable processing bandwidth, *i.e.* single bunch or narrow-band modes can be selected with different filter coefficients.

#### Stability at Synchrotron Light Sources

Requirements for beam stability in synchrotron light sources have driven the development of a number of high-resolution ultra-stable beam position technologies. This began in the 1980s with the Bittner/Biscardi switched receiver concept, [21] developed further in the 1990s by Hinkson and Unser, and now marketed by Bergoz Instrumentation. [22] More recently digital BPMs were developed for global closed-orbit feedback system at the Swiss Light Source [23] which evolved into a product from Instrumentation Technologies [24] employing fast ( $>100$  Msample/sec) data acquisition coupled with digital processing. Recent tests indicate that the AC noise floor approaches  $2 \text{ nm}/\sqrt{\text{Hz}}$  when using capacitive pickup electrodes mounted on an 8 mm high elliptical beam pipe. [25] Long-term stability at the level of 200 nm peak-peak over a 24-hour period has been reported. [26] The Instrumentation Technologies product is used in least 10 recently construct light source facilities. Similar devices have been developed at labs, for example the Swiss Light Source, [27] Advanced Photon Source, [28] SLAC, [29] and NSLS-II. [30]

**Intercepting methods** Fluorescent screens, or flags [31] are very useful and are used extensively in beamlines for both position and beam size and shape measurements. In storage rings the insertion mechanism can be expensive as the beam pipe must be very low impedance when the flag is in the withdrawn position. Monitoring with a CCD camera and frame grabber permits quantitative information quickly. CID cameras may be useful in high radiation applications. Widely used materials are “Chromox6” [32] zinc sulfide and other materials [33]. Graininess can be an issue for small beams. Radiation damage, time response, color, light intensity and stopping power of the flag need to be determined by the designer.

## References

- [1] G. Lambertson, AIP Proc. 153, V. 2 (1985)
- [2] K.Y. Ng and K. Bane, Handbook of Accelerator Physics and Engineering, p. 236
- [3] S. Walston et al, Nucl. Instrum. Meth. A578 (2007)
- [4] S.R. Smith et al, PAC09
- [5] H. Maesaka et al, DIPAC09
- [6] D. Lipka, DIPAC09
- [7] M. Ross et al, PAC03
- [8] C.K. Ng et al, PAC 95
- [9] H.A. Wheeler, Transmission-line properties of a strip on a dielectric sheet on a plane, IEEE Tran. Microwave Theory Tech., vol. MTT-25, p. 631 (1977)
- [10] R. Schafer, AIP Proc. 212 (1989)
- [11] T. Linnekar, CERN-SPS/ARF/78-17
- [12] J. Hinkson and K. Rex, PAC 95
- [13] T. Katsura and S. Shibata, KEK-79-27 (1979)
- [14] R. Johnson, PAC 97
- [15] For example Meggitt Safety Systems, Simi Valley, California; Times Microwave, Wallingford, CT; Ceramaseal, Laurens, SC
- [16] R. Schafer, AIP Proc. 212 (1989)
- [17] D. Cocq and G. Vismara, BIW 98
- [18] V. Schlott et al, DIPAC (2001)
- [19] W. Schappert et al, PAC 03
- [20] R. Kimball, AIP Conf. Proc. 648 (2002)
- [21] R. Biscardi and J. Bittner, PAC89
- [22] Bergoz Instrumentation, St Genis-Pouilly, France
- [23] V. Schlott et al, DIPAC01
- [24] Instrumentation Technologies, Solkan, Slovenia
- [25] G. Decker et al, BIW10
- [26] G. Rehm, EPAC08
- [27] B. Keil et al, EPAC08
- [28] H. Bui et al, BIW08

- [29] E. Medvedko et al, PAC09
- [30] K. Vetter et al, BIW10
- [31] J. Galayda, AIP Conf. Proc 212 (1989)
- [32] Chromox Screens, Morgan Matroc Ltd. Surrey, UK
- [33] G.R. Aiello and M.R. Mills, NIM A346 (1994)

## 7.4.7 Longitudinal Distribution Function, Electrons

*B. Schmidt, DESY*

### 7.4.7.1 Longitudinal diagnostics with coherent radiation

With the availability and growing use of subpicosecond electron bunches, the observation of coherent radiation from these bunches in the infrared and THz regime became a valuable tool to measure the bunch length and longitudinal structure since time domain measurements are difficult and expensive in this regime. Coherent synchrotron radiation (CSR, first observed in 1989 [1]), transition radiation (CTR [2]) and diffraction radiation (CDR [3, 4]) are the most used radiation processes for bunch length diagnostics.

**Principles** The emission process of an extended charge distribution (bunch) can be described by a linear superposition of the electric fields (in time-domain) of the  $N$  individual particles, the wavelength dependent energy density spectrum into a solid angle interval is given by

$$\frac{dU}{d\lambda d\Omega} = \left( \frac{dU}{d\lambda d\Omega} \right)_1 (N + N(N-1) |F(\lambda, \Omega)|^2) \quad (1)$$

with  $F(\lambda, \Omega)$  the (complex) bunch form factor and  $\left( \frac{dU}{d\lambda d\Omega} \right)_1$  the spectrum of a single charge.  $F(\lambda, \Omega)$  is given by the Fourier transform of the normalized particle density distribution  $S_{3D}$  as

$$F(\lambda, \Omega) = \int S_{3D}(\vec{r}) e^{-i\vec{k} \cdot \vec{r}} d\vec{r} \quad (2)$$

where  $\vec{k}$  is the wave vector pointing in the direction of the observer ( $|\vec{k}| = 2\pi/\lambda$ ) and  $\vec{r}$  the position vectors of the individual charges. For highly relativistic bunches ( $\gamma \gg 1$ ), radiation is typically emitted into a narrow cone with opening angle  $\theta \approx 1/\gamma$ . In this case, the transverse size  $\sigma_T$  of the bunch is negligible for wavelengths  $\lambda > \sigma_T/\gamma$  and the form factor is determined by the longitudinal structure of the bunch

$$F(\lambda, \Omega) \approx F_L(\lambda) = \int S(z) e^{-i2\pi z/\lambda} dz \quad (3)$$

alone [5].

**Integrating monitors** Integrating bunch length monitors rely on the fact that the overall emitted coherent radiation power increases with decreasing bunch length. If integrated over a sufficiently large wavelength range, the intensity is largely insensitive to the details of the bunch structure. Figure 1 shows the integrated CTR power as function of the bunch length for a Gaussian and a rectangular bunch of same rms width integrated from  $10 \mu\text{m}$  to  $1000 \mu\text{m}$ . Under realistic conditions, the radiation monitor has a finite spectral coverage (aperture effects, windows, detector response), but the strong correlation between bunch length and coherent intensity still holds. Pyro-electric sensors combine a broad wavelength range with moderate sensitivity; thin crystals with fast readout achieve a sensitivity of about  $10 \text{ nJ}$  for  $100 \text{ ns}$  shaping time. GHz and THz diode detectors have superior sensitivity and time response, but cover a limited wavelength range of typically  $\pm 50\%$  of their center frequency.

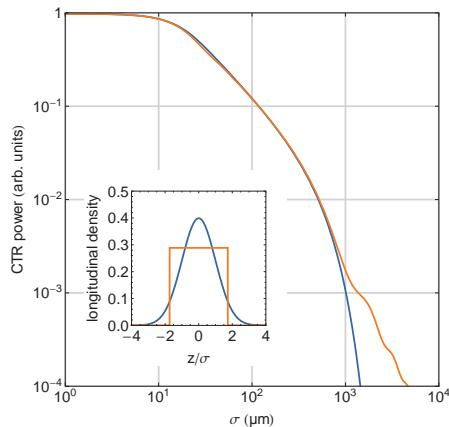


Figure 1: Integrated power of coherent transition radiation as function of bunch length from Gaussian and rectangular bunches of identical rms width. The radiation is integrated from  $10 \mu\text{m}$  to  $1000 \mu\text{m}$ .

**Wavelength resolving monitors** As seen from Eq.(1), spectrally resolved measurements of coherent radiation reveal the absolute value of the (longitudinal) form factor, if the single charge spectrum of the source is known. Concerning the large wavelength range which has to be covered to infer the bunch shape from the spectrum, it is mandatory to use broadband vacuum windows (diamond) and to avoid absorption in humid air in

the optics and spectrometer [7]. Despite the fact that an intensity measurement does not provide any phase information and only the full knowledge of the complex form factor would allow a unique reconstruction of the charge distribution, it has been shown early [6] that also a limited knowledge of  $|F_L|$  is sufficient to reconstruct non-trivial bunch shapes using the Kramers-Kronig relation for computing the phases.

*Interferometric techniques* Interferometric techniques, using Michelson-type interferometers with polarizing beam splitters, have been used to measure the autocorrelation function of the coherent radiation pulses, which is the Fourier transform of the power spectrum. The wavelength coverage of these devices is defined by the characteristics of the detectors and polarizers. Being step-scan devices, they measure the average spectrum of many bunches and slow detectors with broadband flat response function are ideally suited (Golay cells, integrating pyroelectric sensors, bolometers). A typical application can be found in [9].

*Grating based techniques* Reflective gratings can be used as dispersive elements throughout the entire wavelength range from the THz to the visible region. In combination with multichannel detectors, they can be used to set up single-shot spectrometers. Intrinsically, a single grating can not disperse more than one octave in wavelength without mix-up from higher orders. The broadband bunch-radiation needs additional filtering. An elegant solution is to use cascaded reflective blazed gratings as both, dispersive elements and pre-filters for subsequent stages [10]. Based on this technique, a four stage device with parallel read-out can cover a full decade of wavelengths single shot, with two selectable grating sets two decades in wavelength can be used to cope with a wide range of possible bunch profiles. Fig.2 shows a comparison of the expected and measured absolute form factor for a linearly compressed electron bunch at the free electron laser FLASH using such a device [11]. Notice that the measured spectrum extends to shorter wavelengths below the resolution limit of the transverse deflecting RF structure (TDS) used to measure the longitudinal profile. Since the beam has to pass a small  $R_{56}$  between the location of the THz spectrometer and the TDS, the bunch shape can be slightly modified and an exact match of the maxima and minima in the form factor can not be expected.

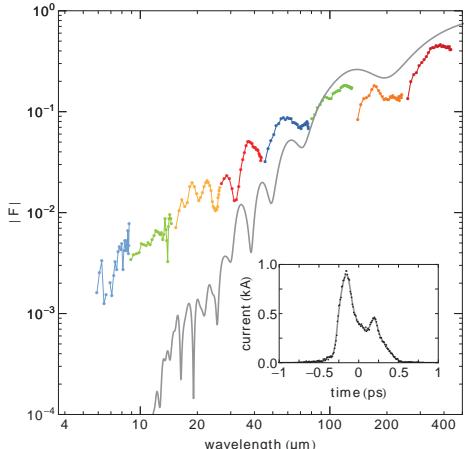


Figure 2: Absolute value of the longitudinal form factor as function of wavelength of a compressed electron bunch at FLASH. Gray line: As derived from the current profile measured with a transverse deflecting RF structure (shown in the inset). Data points : Measured with a four stage single shot THz spectrometer.

## References

- [1] T. Nakazato et al, PRL 63 (1989) 1245
- [2] U. Happek et al, PRL 67 (1991) 2962
- [3] W. Barry, AIP Conf. Proc. 390 (1997) 173
- [4] M. Castellano et al, PR E63 (2001) 056501
- [5] R. Lai and A.J. Sievers, NIM A397 (1997) 221
- [6] R. Lai et al, PR E50 (1994) 4294
- [7] S. Casalbuoni et al, PR ST12 (2009) 030705
- [8] S. Wesch et al, FEL09 Proceedings (2009), WEPC50
- [9] R. Thurman-Keup et al, FERMILAB-PUB-08-115-AD (2008)
- [10] H. Delsim-Hashemi, PhD Thesis Univ. Hamburg (2005) desy-thesis-08-024
- [11] S. Wesch et al, NIM A665 (2011) 40

### 7.4.7.2 Electro-optical bunch length monitors

**Principles** When a relativistic bunch passes within a few millimeters of an electro-optic (EO) crystal, its transient electric field is equivalent to a half-cycle THz pulse impinging on the crystal. The temporal profile of this pulse provides a faithful image of the longitudinal charge distribution in the bunch if the particles are highly relativistic. The transient electric field induces birefringence in the electro-optic crystal. Due to this, a co-propagating linearly polarized laser pulse

acquires an elliptical polarization and thus an orthogonal component which can be detected with high sensitivity using a crossed polarizer-analyzer setup. It directly images the temporal profile of the electric field and thus the longitudinal structure of the bunch. The most common crystals used for EO detection are ZnTe and GaP. The phase retardation of the two polarization components is optimal if the crystal is cut in the (110)-plane and the direction of the electric field is parallel to the [-110] axis. In this case, the relative phase shift for a laser field with wavelength  $\lambda_0$  in a slice of thickness  $dz$  is given by

$$d\Gamma = \frac{2\pi}{\lambda_0}(n_1 - n_2)dz = \frac{2\pi}{\lambda_0}n_0^3 r_{41} E_{THz} dz \quad (4)$$

with  $n_0$  the field-free refractive index and  $r_{41}$  the electro-optic coefficient of the material. The laser and THz pulses travel inside the crystal with different group velocities leading to a phase mismatch for thicker crystals limiting the achievable phase retardation. Together with the transmission coefficient (vacuum to crystal), the effect is described by the geometric response function  $G(\omega, d)$  for the Fourier component with frequency  $\omega$  of the electric field and a crystal of thickness  $d$

$$G(\omega) = \frac{2}{1 + \tilde{n}(\omega)} \frac{1}{d} \int_0^d \exp \left[ i \left( \frac{\omega z}{v_{ph}(\omega)} - \frac{\omega z}{v_g} \right) \right] dz \quad (5)$$

$\tilde{n}(\omega)$  denotes the complex refractive index,  $v_{ph}(\omega)$  the phase velocity at the THz frequency and  $v_g$  the optical group velocity at the laser wavelength. While equation (5) is valid for infinitely short laser pulses it has to be extended to cover finite laser pulse lengths and the dispersive broadening inside the crystal. The product  $G_{EO}(\omega) = G(\omega)r_{41}(\omega)$  defines the *electro-optic response function*  $G_{EO}(\omega)$  for a given laser wavelength  $\lambda_0$  and pulse width. If  $\hat{E}(\omega)$  denotes the Fourier transform of the THz electric field, the total phase retardation (signal) in a crystal of thickness  $d$  is then given by

$$\Gamma(t) = \frac{2d}{\lambda_0} n_0^3 \operatorname{Re} \left( \int_0^\infty e^{-i\omega t} \hat{E}(\omega) G_{EO}(\omega) d\omega \right) \quad (6)$$

Fig.3 shows the response function for GaP crystals of different thicknesses at  $\lambda_0=800$  nm. Using sufficiently thin crystals, field components up to about 7 THz can be detected. The phonon resonances (11 THz in GaP, 5.3 THz in ZnTe) finally

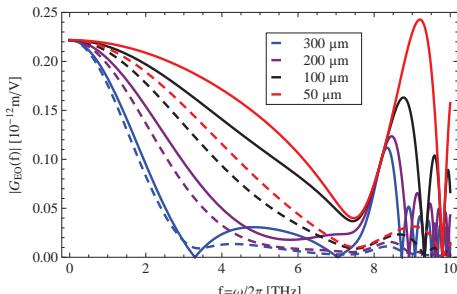


Figure 3: Electro-optic response function of GaP at  $\lambda_0=800$  nm for crystals of 50  $\mu\text{m}$  to 300  $\mu\text{m}$  thickness. Solid lines:  $\sigma_{Laser}=0$  fs, dashed line:  $\sigma_{Laser}=35$  fs.

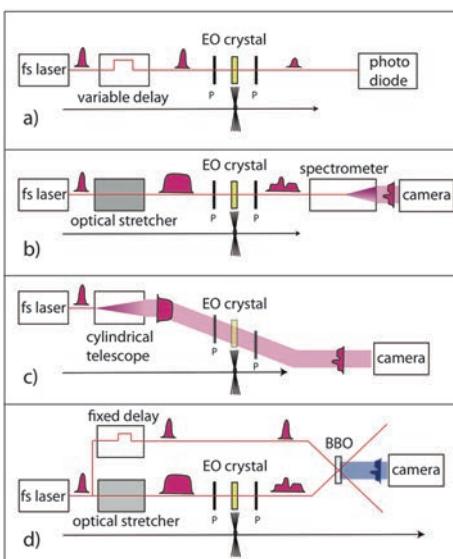


Figure 4: The techniques to de-code the E-field profile from the probing pulses : a) scanning, b) spectral c) spatial and d) temporal decoding.

limit the bunch lengths which can be reliably reconstructed by EO techniques to about 50 fs (rms) for GaP and 100 fs (rms) for ZnTe [1].

**Detection schemes** Common to all EO techniques (Fig.4) is, to detect the polarization change by placing the EO crystal between two crossed polarizers thus measuring the intensity of the field-induced polarization component. Typically additional  $\lambda/4$  and  $\lambda/2$  plates are used to compensate the intrinsic birefringence and to

optimize the S/N ratio. In the first used and technically simplest technique, the temporal profile of the electric field is probed by a comparatively short laser pulse in a step-scan technique (EOS) varying the delay between the laser pulse and the bunch from shot to shot. While EOS in principle has the ultimate resolution, it is restricted to cases where laser and beam rate are perfectly synchronized. At a laser driven plasma wakefield accelerator, a resolutions of 50 fs has been achieved [3]. Typically the time resolution of the EOS method is spoiled by a time jitter larger than the bunch length or shot-to-shot fluctuations of the bunch profile. In this case, single shot techniques are mandatory. For *spectral decoding* [2], a short laser pulse is stretched to a few picoseconds by passing through a dispersive medium generating a temporal wavelength chirp. The time domain modulation is thus decoded in a wavelength domain modulation which can be measured single shot by a spectrometer - camera combination. The method is technically simple but has intrinsic resolution limitations. The intensity modulation of the pulse (the signal) inevitably creates additional frequency components compromising the ideal frequency-time relation of the linear chirp [4]. Signal distortions and resolution limitations become apparent for structures shorter than  $\sqrt{\frac{1}{2}t_c t_0}$ ,  $t_0$  being the initial,  $t_c$  the chirped pulse length. Spectral decoding is ideal for bunch lengths in the ps regime [5]. *Spatial decoding* [6] uses non collinear cross-correlation between THz field and laser pulse. It is free of the above mentioned limitations but requires more complicated optics and imaging techniques close to the beam. Signals widths down to 70 fs (rms) have been reported [7]. *Temporal decoding* finally [8] [9] avoids the frequency mixing problems by sampling the chirped pulse by a second short pulse in a non-collinear optical cross correlator. It requires complex optics including a high power short pulse laser to drive the second harmonic generation in the BBO crystal. Using a 65  $\mu\text{m}$  GaP crystal, a resolution of 60 fs (rms) has been reported [10]. Figure 5 shows the single-shot reconstructed profile of a non-linear compressed electron bunch at the FLASH linac.

**Towards shorter bunches** The drive towards fs bunches has made the advance of electro-optical methods an object of continuing research as, for example in [11].

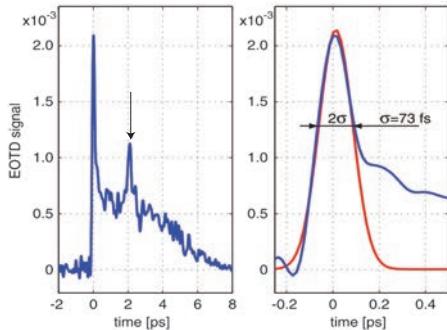


Figure 5: Single shot reconstruction of the bunch shape at the FLASH linac using electro-optic detection in a 100  $\mu\text{m}$  GaP crystal with temporal decoding of the signal [10]. The peak with the arrow is caused by internal reflection in the GaP crystal.

## References

- [1] S. Casalbuoni et al, PR ST11 (2008) 072802
- [2] Z. Jiang, X.-C. Zhang, APL72 (1998) 1945
- [3] J. van Tilborg, BIW08 (2008) 65
- [4] S.P. Jamison et al, Optics Lett. (2006) 1753
- [5] I. Wilke et al, PRL 88 (2002) 124801
- [6] J. Shan et al, Optics Letters 25 (2000) 426
- [7] A. Azima et al, Proc. EPAC06 p71
- [8] S.P. Jamison et al, Opt. Lett. (2003) 1710
- [9] G. Berden et al, PRL 93 (2004) 114802
- [10] B. Steffen et al, PRST AB 12 (2009) 032802
- [11] M.H. Helle et al PRST-AB 15 (2012) 052801

### 7.4.8 Transverse and Longitudinal Emittance Measurements

*J.T. Seeman, SLAC*

Emittance measurements require the use of a profile monitor. Among the various possible monitors that intercept the beam are fluorescent screens [1], wire scanners [2], optical transition radiators [3], collimator edges with a downstream current monitor (or spray detector) and beam-beam scans [4]. Non intercepting monitors include scattering laser light (See Sec.7.4.12 and 7.4.13), synchrotron radiation [5], gas ionization projection monitors [6], profiles from luminosity scans [7] and profiles from particle physics detectors [8]. Coherent radiation spectra from intercepting (i.e. transition radiators) and non-intercepting sources (i.e. SR) can also be used (Sec.7.4.13, 7.4.10).

**Longitudinal distribution function and emittance** The energy spread,  $\delta = \sigma_E/E$  can be

measured in a dispersive beam line where the dispersion  $D$  is large enough that the energy spread contribution to the beam size dominates, i.e.  $(D\delta)^2 \gg \epsilon\beta$ . The measurement is then

$$\delta = \sigma/D$$

with  $\sigma$  being the transverse beam size in the dispersive plane. RMS values for  $\delta$  and  $\sigma$  are usually used. As a typical example, for  $D \sim 0.5$  m and a beam size measured to 500  $\mu$ m, the energy spread resolution is about 0.1%.

A method to determine the combined energy spreads of the beams in an  $e^+e^-$  collider is to scan over a narrow excitation resonance and remove quadratically the natural energy spread of the particle resonance. An example is operating near the Y(2S) resonance where the full resonance width is 32 keV out of a resonance energy of 10 GeV [9].

The bunch length  $\sigma_z$  can be measured in several ways: (i) streak camera (Sec.7.4.11) or fast photodiode; (ii) spectrometer; (iii) resonant cavity; or (iv) coherent radiation spectrum (Sec.7.4.13).

In (i) and (iv) the beam is first passed through a radiator to generate photons, e.g. SR from a magnet, transition radiation from a foil or Cerenkov radiation from a thin quartz crystal. Streak cameras can have resolutions down to 200 fs and fast diodes to 10 ps.

In (ii) a bunch is made to pass through an RF cavity section [10] with the longitudinal center of the beam timed to cross at the zero of the longitudinal rf voltage. Thus, particles ahead of center are accelerated (decelerated) and tail decelerated (accelerated) in proportion to distance from the bunch center. Downstream the beam is sent into a spectrometer where the dispersion induced component of the beam size dominates the betatron size.

$$\sigma = D(E_{rf}/E_0) \sin[2\pi(\sigma_z/\lambda_{rf})]$$

$\sigma$  being the measured beam size,  $E_{rf}$  being the peak cavity voltage and  $E_0$  being the central beam energy. This can be solved for  $\sigma_z$ .

A similar technique is to pass the bunch through a transverse deflecting RF cavity with the head deflected oppositely from the tail and then observe the width of the bunch directly downstream. This deflection technique has been used to measure various bunch parameters: lengths, energy spreads, head-tail energy differences, slice emittances and energy spreads [11, 12]. Bunch

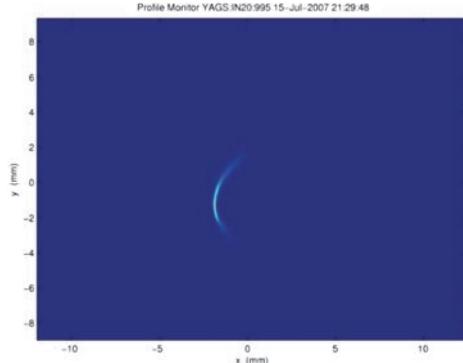


Figure 1: Profile of the particle energy (horizontal) versus longitudinal particle position (vertical) by passing a bunch through a vertical transverse deflection cavity in the LCLS injector observed on a downstream spectrometer screen [12].

lengths have been measured by this method to about 5 fs. An example is shown in Fig.1 (See also 7.4.13)

In (iii) an unpowered resonant cavity [13] is employed. The beam passes a ceramic gap and a portion of the field energy thus radiated outward is captured by an off axis resonant cavity. The radiated power  $P$  is related to the bunch length and current  $I_0$ ,

$$P \propto I_0^2 e^{-\left(\frac{\omega z}{c}\right)^2}$$

$\omega$  being the cavity frequency. The power collected by the cavity can be measured with a high frequency power meter and the bunch length determined. The cavity frequency is matched to the expected bunch length. Several cavities may be employed simultaneously to obtain information about the bunch spectrum. Calibration of such monitors may be difficult.

**Transverse distribution function and emittance** Transverse emittances can be measured using the lattice optics of a transport line and a profile monitor or monitors located where the betatron size of the beam dominates, i.e.  $\epsilon\beta \gg (D\delta)^2$ . In general we wish to determine the  $\epsilon$ ,  $\beta$ ,  $\gamma$ , and  $\alpha$  of the incoming beam at, say position 1 using a monitor at position 2. The Courant-Snyder parameters at 1 can be transformed to 2 in the transport line using **R** TRANSPORT matrix elements (Secs.2.1.1, 2.2.1). Thus, the downstream beam size  $\sigma_2$  can be written using the upstream parameters as

$$\sigma_2^2 = \epsilon\beta_2 = \epsilon(R_{11}^2\beta_1 - 2R_{11}R_{12}\alpha_1 + R_{12}^2\gamma_1)$$

The beam size downstream is measured with three non-degenerate sets of  $R_{11}$  and  $R_{12}$  between 1 and 2, resulting in three equations in the three unknowns. There are many possible arrangements for the position of the monitors(s) in the transport line. One of the simplest is the quadrupole-drift-monitor arrangement, using quadrupole strength to vary the  $\mathbf{R}$  matrix [14]. Alternatively, there may be many quadrupoles between the varied quadrupole and the monitor. Another common arrangement is for three (or more) profile monitors to be spaced along a transport line, requiring no change in quadrupole strength. This is good for high power beams and delicately tuned beams. The  $\mathbf{R}$  matrix changes naturally along the line. In the example for the quadrupole – drift – monitor method, we have  $R_{11} = 1 + Lk$  and  $R_{12} = L$  where  $L$  is the drift length from the quadrupole to monitor and  $k$  the quadrupole strength. Then

$$\sigma_2^2 = \varepsilon[(1 + Lk)^2\beta_1 - 2(1 + Lk)L\alpha_1 + L^2\gamma_1]$$

displaying the characteristic quadratic change in beam size squared with  $k$ . The quadratic term of the parabola gives the term  $\varepsilon(Lk)^2\beta_1$  directly. Of course these same stratagems can be used in closed orbit machines as well if one has measurements of the beam profile by one of the nonintercepting means turn-by-turn.

Particularly in linacs, many effects will cause the phase space distribution to deviate from Gaussian (Sec.4.3). These non-Gaussian distributions can be fit in several ways including skew moments (3rd order), kurtosis moments (4th order), an asymmetric Gaussian with a different  $\sigma$  for the left and the right sides [15], or a fit to wakefield tails of the beam.

As an example of such a distribution, transverse wakefields in the linac may cause various particles to be displaced transversely, depending on their  $z$  position in the bunch, resulting in a transverse tail (beam break-up, Sec.2.4). This tail is then transported along the linac with a betatron phase advance. This distribution can be calculated or fit using measured beam profiles. A fitting method [16] uses the expected offsets to allow longitudinal slices of the bunch to be displaced in a monotonic exponential offset from head to tail. The slice offsets are given by a function of betatron phase and  $z$ ,

$$\frac{x_0(\phi_i, z)}{\sigma_i} = U(z_0 - z) \left[ e^{(z_0 - z) \frac{\tau}{\sigma_z}} - 1 \right] \cos(\phi_i + \phi_0)$$

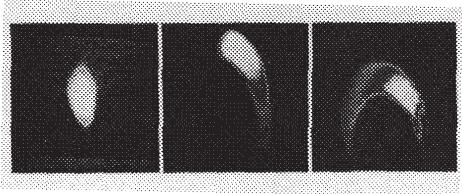


Figure 2: Observed transverse wakefield tails in the SLAC 3 km linac. The left image is for a well steered beam. The center image is for an oscillation of about 0.5 mm and the right image for 1.0 mm. The bunch intensity is  $2 \times 10^{10}$  electrons and the core sizes are about 120  $\mu\text{m}$  in both x and y.

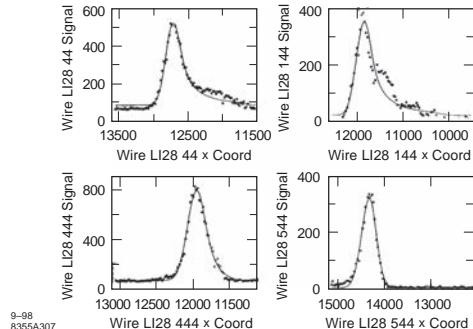


Figure 3: Fit profile to transverse wakefield tails.

where  $z$  is distance along the beam line,  $U$  the unit step,  $\sigma_i$  the width of slice “ $i$ ”,  $z_0$  the starting position of the exponential offset,  $\tau$  the offset parameter of the tail and  $\phi$  the betatron phase advance, giving for the transverse distribution of each slice,

$$\rho(x, \phi_i, z) = \frac{e^{-[(x-x_0)^2/2\sigma_i^2]}}{\sqrt{(2\pi)\sigma_i}}$$

Integrating over the bunch length gives the final distribution. The parameters  $\phi_0$ ,  $\tau$ , and  $z_0$  are chosen to fit the beam shape over a short section of the linac. For example, an observed wakefield enlarged beam is shown in Fig.2 and Fig.3 displays the non-Gaussian fits to the 4 profile (wire) monitor data [16]. The monitors are located sequentially at 0, 22, 90 and 112 degrees (upper left to lower right) in betatron phase. One can see the near exponential growth of the tail without curvature as the beam passes down the linac. The tail lines up with the head in the lower right panel.

## References

- [1] M. Ross, SLAC-PUB-3640, SLAC (1985)
- [2] M. Ross et al, PAC91, p.1201
- [3] R. Riorito and D. Rule, AIP Proc. 319 (1993)
- [4] M. Venturini et al, PAC01, p.3573
- [5] A. Fisher, LHC Performance Note-014 (2009)
- [6] V. Lepeltier et al, EPAC96, p.1693
- [7] W. Kozanecki et al, PAC05, p.1874
- [8] B. Viaud et al, EPAC06, p.598
- [9] Y. Cai et al, PRSTAB 12, 061002 (2009)
- [10] J. Seeman, AIP Proc. 249 (1992) p.49
- [11] R. Akre et al, EPAC02, p.1882
- [12] J. Frisch et al, BIW 08, p.17
- [13] E. Babenko et al, PAC 93, p.2423
- [14] M. Ross et al, PAC 87, p.725
- [15] W. Spence, F.J. Decker, M. Woodley, PAC93, p.3576
- [16] M. Avilov et al, High En. Acc. Conf. (1992) p.882

### 7.4.9 Transverse Beam Profile Measurements with Transition and Diffraction Radiation

*P. Evtushenko, JLab*

Transition radiation (TR) is generated when charged particle passes through an interface of two media with different dielectric constant (see Sec. 3.1.6). The most common arrangement used in beam diagnostics is a metal foil inserted in to the beam i.e. an interface of vacuum and metal. The spectrum of TR is very broad and extends to the plasma frequency of the metal at the high frequency end. Since the plasma frequency of a metal is on the order of  $10^{16}$  Hz, the yield of TR is high up to a few tens of nm. The part of the TR lying in the optical range is commonly known as optical transition radiation (OTR) and is used for transverse beam profile measurements. When the bunch length is much longer than the optical wavelength the OTR is longitudinally incoherent and its local intensity is linearly proportional to the local beam current density. This makes OTR ideal for transverse beam profile measurements. OTR is mostly used for profile measurements of electron beams, however has been used successfully for proton beam measurements with very high energy [1]. For electron beam measurements OTR has been used in a range from 10 keV [2] up to 30 GeV [3].

The angular distribution (AD) of the TR plays important role in the design of the beam

diagnostics. For a single particle it can be derived by analogy with Bremsstrahlung radiation considering TR as the interaction of the charged particle with its image charge particle. Then the distribution of the transition radiation is written as [4]:

$$\frac{d^2W}{d\omega d\Omega} = \frac{q^2}{4\pi^2 c} \left| \sum \left( \frac{\vec{\beta}_f \times \vec{s}}{1 - \vec{s} \cdot \vec{\beta}_f} - \frac{\vec{\beta}_i \times \vec{s}}{1 - \vec{s} \cdot \vec{\beta}_i} \right) \right|^2, \quad (1)$$

where  $W$  is the radiated energy,  $\omega = 2\pi f$  is the angular frequency of the radiation,  $\Omega$  is the solid angle,  $q$  is charge of the particle,  $c$  is speed of light,  $\vec{s} = \vec{k}/|k|$  is unit wave vector,  $\vec{\beta}_f$  and  $\vec{\beta}_i$  are the velocities of the particle normalized to the speed of light after and before the interaction. In the Eq.(1) the sum is taken over the particle and its image current imaginary particle. The Eq.(1) is an approximation valid for perfectly conducting and infinitely large radiator in the far field i.e. in the Fraunhofer approximation. The most comprehensive way to calculate the AD of the TR is to use vector diffraction calculations [5].

One of the key parameters of any transverse diagnostic is the spatial resolution. It was shown [6]-[8] that the OTR resolution is diffraction limited and is essentially independent of beam energy. The spatial resolution is described in terms of the point spread function (PSF); the radiation intensity distribution in the image plane obtained from infinitely small beam. The distribution of the radiation intensity from a beam is a 2D convolution of the beam profile and the PSF. Thus, for proper measurements the optical system must be designed to have a PSF smaller than the required resolution. When the charged particle is ultrarelativistic, the Fraunhofer approximation is valid and the size of the OTR radiator is much bigger than the size of the OTR source,  $\sim \gamma\lambda$ , the OTR PSF for wavelength  $\lambda$  can be written in a simplified form [7]:

$$I_{OTR} \propto \left( \int_0^{\theta_{lens}} J_1(2\pi r \theta/\lambda M) \frac{\theta^2(\sqrt{1-\theta^2}+1)}{\theta^2 + (\gamma\beta)^{-2}} d\theta \right)^2, \quad (2)$$

where  $M$  is the imaging lens magnification,  $\theta_{lens}$  is its angular acceptance,  $r$  is the transverse coordinate on the image plane measured from the center of the PSF. If the above-mentioned conditions are not met the OTR PSF still can be calculated numerically [5, 7]. FWHM of the OTR PSF is about three times larger than the PSF of ideal point source  $I \propto (J_1(kr\theta_{lens})/(kr\theta_{lens}))^2$ .

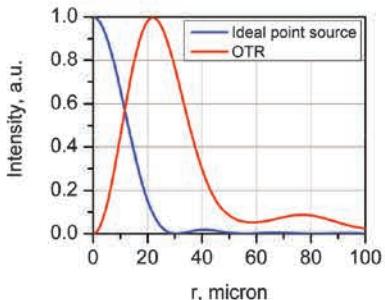


Figure 1: PSF of OTR and ideal point source.

This is due to the difference of angular distribution of the TR and an isotropic point source. In Fig.1 the PSF of OTR and an ideal point source are shown for  $\lambda=500$  nm and  $\theta_{lens}=0.01$ . Systems with a resolution of about 10  $\mu\text{m}$  used for beam size measurements in the range of 50  $\mu\text{m}$  and larger are typical [9, 10]. An OTR based system with a resolution of about one micron was built and used for measurements of a beam with RMS size about 5 micron [11].

Diffraction radiation (DR) has the same nature as TR. It is generated by the fast time varying image currents induced in a radiator surface by the Coulomb field of charged particle. Here the charged particle does not cross the surface of the radiator but only needs to pass close enough to the edge of the radiator. Following demonstrates the required distance magnitude. The amplitude of a Fourier component of a charged particle Coulomb field with longitudinal velocity  $v$  is given by [12]  $E_{r\omega} = \frac{q\alpha}{\pi v} K_1(\alpha r)$ , where  $q$  is the charge of the particle,  $\alpha = 2\pi(\lambda\gamma\beta)^{-1}$  and  $K_1$  is a first-order modified Bessel function of the second kind. At the radiator surface, the field of the radiation has the same amplitude as the Coulomb field of the particle. For the radiation to have an amplitude comparable with the TR  $\alpha r$  needs to be smaller than unity or  $r < \gamma\lambda/2\pi$ . The characteristic size  $\gamma\lambda/2\pi$  is referred to as the size of the source for both TR and DR.

While OTR beam size measurements have been routinely used, those by DR are still in the R&D phase. As opposed to OTR DR has the potential to provide non-intercepting measurements for high average current beam.

To make DR radiation practically useful for measurements in the optical wavelength range, the beam energy and the integrated charge of the

beam needs to be sufficiently high. Good signal to noise ratio was demonstrated at  $\gamma \sim 1300$  and the integrated charge of 20 nC [13].

Two different approaches are considered for transverse beam size measurements via ODR. One is based on the dependency of the ODR AD on the beam size referred to as far-field measurements [14]-[16]. The second is based on imaging of the ODR radiator surface - nearfield measurements [14, 17, 18]. In the first case it was shown that the AD of the ODR generated by a beam passing through a circular aperture or a slit exhibits diffraction fringes. The visibility of the fringes depends on the beam size. The situation is complicated by the fact that the beam divergence and energy spread also affect the visibility of the fringes.

For near-field measurements the observed distribution is a convolution of the beam distribution and the PSF of the ODR. Thus to reconstruct the beam distribution a 2D deconvolution is required. To do this the ODR PSF needs to be verified experimentally and the measured near-field ODR pattern should contain no artifacts introduced by sources others than the DR. In both cases it was demonstrated that measured AD or the near-field pattern size depend on the beam size as predicted by theory. It was found that the visible SR and edge radiation presents significant source of background, elimination of which is one of the present experimental efforts directions.

## References

- [1] V.E. Scarpine et al, PAC07, p.2639
- [2] R.B. Fiorito et al, PAC07, p.4006
- [3] P. Catravas et al, PAC99, p.2111
- [4] J.D. Jackson, Classical Electrodynamics (1975) 703
- [5] A.G. Shkvarunets et al, PRST AB 11 (2008) 012801
- [6] V.A. Lebedev, NIM A372 (1996) 344
- [7] M. Castellano et al, PRST AB 1 (1998) 062801
- [8] X. Artru et al, NIM B145 (1998) 160
- [9] J.-C. Denard et al, PAC97, p.2198
- [10] A. Cianchi et al, EPAC04, p.2619
- [11] M. Ross et al, SLAC-PUB-9280 (2002)
- [12] M.L. Ter-Mikaelyan, High Energy Electromagnetic Processes in Condensed Media John Wiley & Sons Inc (1972)
- [13] E. Chiadroni et al, DIPAC09, p.151
- [14] R.B. Fiorito et al, AIP 472 (1998) 725
- [15] M. Castellano, NIM A394 (1997) 275

- [16] P. Karataev et al, PRL 93 (2004) 244802
- [17] A.H. Lumpkin et al, PRST AB 10 (2007) 022802
- [18] P. Evtushenko et al, BIW08, p.133

#### 7.4.10 X-Ray Beam Size Monitor

*J.P. Alexander, D.P. Peterson Cornell*

Knowledge of beam emittance is important for optimal accelerator operation and may be derived from measurements of the transverse beam size ( $\sigma_x$ ,  $\sigma_y$ ) at a given location. We discuss here the characteristics and performance of the x-ray beam size monitor (xBSM) at the Cornell University  $e^+e^-$  accelerator. The xBSM measures vertical beam sizes ( $\sigma_y$ ) of order  $10 \mu\text{m}$  by imaging  $\sim 2$  keV synchrotron radiation photons onto a one-dimensional photodiode array. Instrumentation in the evacuated x-ray beam line includes upstream interchangeable optical elements (slit, coded aperture, and Fresnel zone plate), an optional monochromator, and an InGaAs photodiode-array detector. The readout is a beam-synchronized FADC that is capable of parallel measurement of consecutive bunches with 4 ns separation. Single turn measurements are fit to characteristic image shapes to extract beam sizes independent of position variations. The turn-averaged beam size provides feedback for low-emittance tuning.

##### 7.4.10.1 X-ray source and optics

The x-ray source point is a standard bend magnet in the storage ring. In the positron (electron) line, the optical elements are located 4.36 m (4.23 m) from the source point and vertically focus the x-ray beam onto the photodiode array situated 10.19 m (10.67 m) further downstream; the magnifications in the two lines are 2.34 and (2.52).

Three optical elements are available for 2 GeV stored beam operation: a vertically limiting slit, a Fresnel Zone Plate (FZP), and a Coded Aperture (CA). These elements reside in the storage ring vacuum and can be selected and aligned remotely to meet the requirements of various measurements. At 2 GeV, the typical power load on the optical element is of order 1 mW/mA; the optical elements are in contact with actively cooled copper supports to remove this heat.

The FZP and CA are manufactured [1] on a common silicon substrate. The patterns are etched into  $0.7 \mu\text{m}$  Au and are supported by a  $2.5 \mu\text{m}$  thick Si membrane. The FZP pattern has 120

transmitting rings in a diameter of  $1200 \mu\text{m}$ . The CA pattern has 8 transmitting slits of varying gap size with total dimensions  $310 \mu\text{m}$  in the vertical (imaging) direction, and  $1200 \mu\text{m}$  in the horizontal.

##### 7.4.10.2 Detector

The detector is a vertical array of 32 InGaAs diodes with pitch  $50 \mu\text{m}$  and horizontal width  $400 \mu\text{m}$ . The InGaAs layer is  $3.5 \mu\text{m}$  thick, which absorbs 73% of photons at 2.5 keV; there is a  $160 \text{ nm } \text{Si}_3\text{N}_4$  passivation layer. The time response of the detector is sub-nanosecond.

The detector resides in an enclosure held at a rough vacuum ( $< 0.5$  torr) to reduce absorption, isolated from the storage ring volume by a thin ( $4 \mu\text{m}$ ) diamond window [2], which transmits 76% of x-rays at 2.5 keV.

Readout of the detector is done in two ways. For alignment purposes and measurements which do not require isolation of bunches, the x-ray image can be scanned by moving a single diode through the image while recording the diode current. The resulting “slow scan” integrates many turns of the beam, thus removing statistical fluctuations but not resolving turn-by-turn fluctuations.

In a “fast scan” pulse heights from the 32 elements of the diode array are simultaneously digitized, synchronous with the bunch crossing. Electronics used for measurements described here allow measurements with 14 ns repetition rate; equivalent performance has been achieved with faster electronics and 4 ns bunch spacing.

##### 7.4.10.3 Beam size measurements

Images of the 2 GeV beam, derived from fast scans of the 32 channel detector, are shown in Fig.1 for the three optical elements. The beam size was similar for the three cases.

The vertically limiting slit (Fig.1, Left) operates as a pin-hole lens. It is largely insensitive to the x-ray wavelength within the synchrotron radiation spectrum. The slit has been adjusted to be about  $45 \mu\text{m}$  in height which gives the minimum image width; a smaller slit height would cause the image to broaden due to diffraction while a larger slit height would cause the image to broaden due to transmission. The pin-hole image is the convolution of the magnified beam profile and the point spread function [3] of the pinhole. The latter has been calculated and can be approximated by a Gaussian beam profile of  $21 \mu\text{m}$  rms.

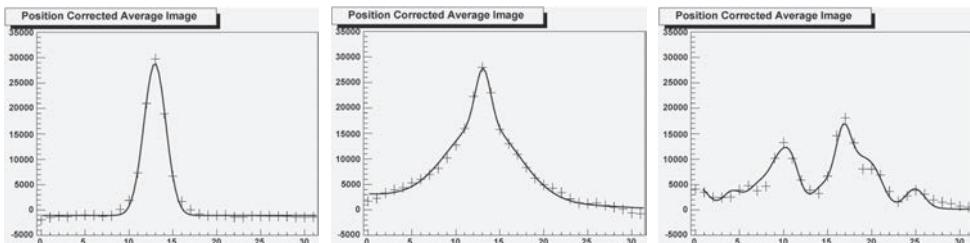


Figure 1: Three images of the beam, taken with three different optical elements for constant beam current 0.5 mA/bunch and vertical beam size 15 microns. The vertical axis is illumination intensity at the detector, with full scale corresponding to 130 photons per turn per bin; the horizontal axis is vertical position in photodiode number. Left: Pinhole; Middle: Fresnel Zone Plate; Right: Coded Aperture. The pinhole and FZP are imaging devices, while the CA produces an illumination pattern that requires additional interpretation to extract the beam size. The intrinsically chromatic FZP exhibits a broad feature due to misfocusing out-of-bandwidth x-rays.

Subtracting in quadrature the  $21\text{ }\mu\text{m}$  spread from the raw beam size measurement yields the corrected beam size. For larger beam sizes, the simplicity of the slit optics is attractive and CesrTA uses it for emittance measurements of all but the smallest beams [4]. The signal amplitude, integrated over the image in 2 GeV beam energy conditions is approximately 500 2.5 keV photons per 1 mA bunch, as determined from the statistical fluctuations of the turn-by-turn signal yields. The distribution shown in Fig.1 is a sum of images taken of the same bunch on 4096 successive turns. To remove the effects of turn-to-turn motion of the bunch centroid, center-corrected individual turn images are summed before fitting to the image shape.

The image of the Fresnel Zone Plate (Fig.1, Middle) is a diffraction pattern and highly sensitive to the x-ray wave length. We have the option to deploy a multilayer mirror monochromator whose  $\sim 1.6\%$  transmission bandwidth is chosen to encompass the  $\sim 1\%$  bandpass of the FZP. The bandpass limitations, however, reduce the signal intensity to a few photons per bunch passing, implying the need for multipass averaging. In cases where multipass averaging is not desirable, the FZP may be operated without the monochromator, yielding the “blurred” image of Fig.1. Because of the filtering of very low energy x-rays by the diamond vacuum window, and the exponential suppression of high energy x-rays inherent in the synchrotron radiation spectrum, the incident x-rays have an effective natural bandwidth of order 20%, enabling one to find a relatively narrow central peak in the image. Thus the image shown

in Fig.1 exhibits both a narrow central peak and a broad underlying distribution of out-of-focus (off-energy) x-rays. The out-of-focus contribution is relatively insensitive to the beam size, allowing the central peak to be unambiguously extracted from a fit to the image. The FZP can be used to provide useful beam size measurements to the smallest beam size.

The illumination pattern of the Coded Aperture (Fig.1, Right) is a combination of transmission and diffraction resulting from the 8 slits which range in transmission gap size from  $10\text{ }\mu\text{m}$  to  $40\text{ }\mu\text{m}$ . The CA does not produce an image in the usual sense, but the illumination pattern can be interpreted in terms of a calculated shape. The latter has a predictable variation with beam size which allows a straightforward fit of the pattern to three variables: beam size, intensity, and position. As in the case of the vertically limiting slit, and in contrast to the FZP, the illumination pattern is relatively insensitive to variations in the wavelength. In contrast to both of the other optics, the CA illumination pattern, in this implementation, fills the available detector. Thus, an independent measure of the background is not possible in a single turn measurement, resulting in a systematic uncertainty described below. The CA is discussed in further detail elsewhere [5].

To understand the relative advantages of the slit, CA, and FZP, data was analyzed for similar beam conditions for the three optics elements. In the case of small beam size, data sets were taken with  $15\text{ }\mu\text{m}$  beam size, 0.5 ma beam current, and negligible turn-by-turn beam motion. Turn-by-turn variations of the beam size measurements

were used to estimate the precision for each of the three optics:

(a) *Slit*. Small beam size measurements using the slit are limited by the need to remove the effect of the slit width (the point-spread function), but improved by its relatively high transmission. While the single-turn, raw beam size measurement has an uncertainty of  $\pm 2.0 \mu\text{m}$ , the corrected beam size has an uncertainty of  $\pm 3.4 \mu\text{m}$ . The single-turn uncertainty grows to  $\pm 4.6 \mu\text{m}$  for a beam size of  $10 \mu\text{m}$ .

(b) *FZP*. Measurements using the FZP are limited by the small fraction of the signal that is in-focus. For a beam size of  $15 \mu\text{m}$ , the single-turn measurements have an uncertainty of  $\pm 3.4 \mu\text{m}$ . The only correction due to instrumentation is for the finite detector pixel size, which contributes  $6 \mu\text{m}$  in quadrature to the beam size measured at the source. For a beam size of  $8 \mu\text{m}$ , the single-turn uncertainty is  $\pm 4 \mu\text{m}$ .

(c) *CA*. The CA has multiple slits and thus transmits more photons, resulting in a statistically improved measurement. As the individual slits of the CA have sizes on the order of the height of the single slit optic, features of the illumination pattern have intrinsic widths which limit the resolution at small beam size and are incorporated into the fit. The typical intrinsic feature size is equivalent to the broadening of the features due to a beam size of  $\sim 15 \mu\text{m}$ . For  $15 \mu\text{m}$  beam size, the single-turn measurement has a variation of  $\pm 2.3 \mu\text{m}$ . For a beam size of  $8 \mu\text{m}$ , the single-turn uncertainty increases to  $\pm 3.5 \mu\text{m}$ , which is similar to the resolution when using the FZP.

For  $N$  passes of the beam, the measurement uncertainties scale as  $1/\sqrt{N}$  when turn-to-turn motion of the bunch centroid is corrected for.

In the case of large beam size, measurements with the FZP are limited when the width of the in-focus signal becomes comparable to the out-of-focus signal at a beam size of  $\sim 45 \mu\text{m}$ . Measurements with the CA are limited by the absence of an independent measurement of the background. For a reasonable estimate of the uncertainty of the background, the systematic uncertainty of the beam size measurement increases above 20% for beam sizes above  $\sim 60 \mu\text{m}$ . Measurements with the slit are stable up to beam sizes of  $\sim 110 \mu\text{m}$  with  $3\sigma$  containment of the image on the detector.

#### 7.4.10.4 Summary

The xBSM routinely provides beam size measurement support for low emittance tuning and dynamic emittance growth at CesrTA. The vertically-limiting-slit, Fresnel Zone Plate, and Coded Aperture allow measurements of beam sizes between  $8 \mu\text{m}$  and  $110 \mu\text{m}$ , with accuracies for small beam sizes at the level of  $2\text{-}4 \mu\text{m}$  in a single-pass measurement.

### References

- [1] Applied Nanotools Inc., Edmonton, Alberta, Canada; <http://www.appliednt.com/>
- [2] Diamond Materials GmbH, Freiburg, Germany; <http://www.diamond-materials.com/EN/index.htm>
- [3] E. Hecht, Optics, 4th Ed., Addison & Wesley 2001
- [4] D. Sagan et al, Proceedings of the 2010 International Particle Accelerator Conference, Kyoto, Japan; <http://epaper.kek.jp/IPAC10/papers/thpe046.pdf>
- [5] J.W. Flanagan et al, Proceedings of the 2010 International Particle Accelerator Conference, Kyoto, Japan; <http://epaper.kek.jp/IPAC10/papers/mope007.pdf>

#### 7.4.11 Streak Cameras

*J.M. Byrd, LBNL*

##### 7.4.11.1 Principle of operation

A streak camera (SC) is an electro-optic detector for measuring the temporal distribution of an optical pulse on a picosecond time scale. In accelerator applications [1, 2, 3], it is typically used for measuring the bunch length from the pulse of visible or near visible radiation emitted by a particle beam. Forms of radiation include synchrotron, edge, undulator, Cerenkov, and transition radiation. SCs are also used for measuring the relative timing of beam pulses or a beam and optical laser pulse. SCs can also be configured as a gated camera.

Shown in Fig.1 is a schematic layout of a SC [4]. An optical radiation pulse from the beam is focussed through an entrance slit to the camera and incident on a photocathode. Photoelectrons are emitted with the same temporal distribution as the optical pulse and are accelerated as quickly as possible, typically through a DC voltage. The photoelectrons pass through a time-dependent vertical deflecting, or streaking, electric field which oppositely deflects the head and

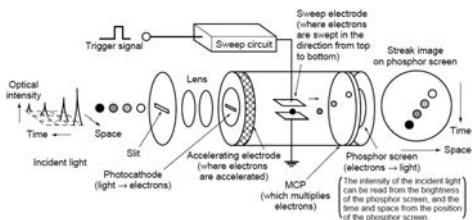


Figure 1: Schematic of a streak camera.

tail of the pulse. The electron pulse is then focussed onto an imaging screen, typically a multichannel plate followed by a phosphor screen. The phosphor screen image is captured with a CCD camera and digitally processed. A multichannel plate (MCP) can be used to increase the signal gain.

If the vertical deflection is large compared with the undeflected vertical beam size at the screen, the temporal distribution of the optical pulse can be inferred from the vertical distribution imaged on the screen. Note that the horizontal dimension still contains information about the horizontal size of the input optical pulse.

Optical SCs are available from several commercial vendors [5] with typical time resolution of 1-2 psec for input optical pulses in the near infrared to near ultraviolet wavelength range. Specialty SCs can be found with time resolution as low as 200 fsec for optical pulses. Models are available with photocathodes suitable for x-ray input pulses as well. Additional features are available such as synchroscan and dual-axis sweep. In the case of synchroscan, the vertical deflection is driven by a radiofrequency signal instead of a single ramped voltage pulse, allowing measurement of periodic optical signals with repetition rates of up to several hundred MHz, such as those from a storage ring or mode-locked laser. In the case of a dual-axis sweep, a second set of deflection plates are added which provide a slow horizontal sweep to the photoelectron beam in addition to the fast vertical sweep allowing measurement of repetitive input signals on a single image.

#### 7.4.11.2 Temporal resolution

The temporal resolution of the SC is limited by a number of factors that have been described in several reviews [6, 7]. The primary contributions are described below.

**Photoelectron velocity spread** Electrons emitted from the photocathode via the photoelectric effect have a spread in longitudinal velocities which leads to temporal broadening of the electron pulse up to point that the electrons reach the deflecting plates. The spread in velocities depends strongly on the incident photon energy and the work function of the photocathode. The effect of the temporal dispersion is greatest during the acceleration of the electrons from the photocathode through the anode. There is further time-of-flight dispersion from the anode to the deflecting voltage. This effect is minimized by accelerating the electrons as quickly as possible from the cathode by maximizing the anode voltage and minimizing the distance from the anode to the deflecting electrode.

Another considerable contribution to broadening of the pulse is photoelectron velocity spread induced by electron-electron Coulombic repulsion (i.e. space charge.) The primary effect is the acceleration and deacceleration of the head and tail of the bunch, respectively, leading to further temporal broadening of the electron pulse. For a given photoelectron intensity, the space charge effects are minimized by accelerating the electrons as quickly as possible from the cathode. To minimize space charge effects and achieve the best temporal resolution, the input light intensity is reduced to smallest level for which an image can still be recorded with an acceptable signal level.

**Spatial imaging resolution** The image of the unstreaked photoelectron beam presents the ultimate limit to the temporal resolution. To minimize the image size, the optical image is focussed on the photocathode via the input optics with the minimum vertical slit and the photoelectron beam is magnetically focussed onto the image screen by a solenoid. When the electron beam is streaked, the vertical size of the imaged electron beam is proportional to the slope of the deflecting voltage. For a pulsed deflecting system, resolution is increased by reducing the rise time and increasing the voltage of the deflecting pulse. For a synchroscan system, temporal resolution is increased by increasing the frequency and voltage of the deflecting signal.

**Optical dispersion** The input optical pulse can be lengthened by dispersion in refractive optics. To minimize this, one must use either reflective optics or an optical bandpass filter.

**Timing jitter** To maximize the dynamic range of the swept image, it is best to adjust the timing of the sweep voltage such that the temporal center of the electron pulse receives no deflection. If the signal intensity is such that multiple pulses are required to accumulate an image, jitter in the relative timing of the electron pulse and sweep voltage broaden the apparent measured pulse length. For a high repetition rate of input pulses, the synchroscan feature is useful for accumulating images because the synchroscan frequency can be phase-locked to the repetition rate of the input pulse, typically the accelerator master clock or its subharmonic, with a significant reduction in jitter. Note that synchrotron oscillations of the particle create and additional timing jitter which must be corrected. For applications with a pulsed sweep voltage, shot-to-shot timing jitter can cause significant broadening of the measured pulse length.

**Signal sensitivity** The observed intensity of the image primarily depends on the QE of the photocathode which is typically about 10% at optical wavelengths. Other factors include the transmission efficiency of the SC tube, gain of the phosphor screen, and sensitivity of the CCD. The use of an MCP allows sensitivity to single photoelectrons. Several SCs have implemented gating of the anode grid voltage and the MCP to reduce the accumulation of noise signal except when beam signal is present.

#### 7.4.11.3 Applications to measurement of beam dynamics

Streak cameras have been used for a wide variety of measurements of accelerator beam dynamics beyond measurement of the pulse length. A few examples illustrating the ability of streak cameras to capture transient behavior are given below.

**Transient longitudinal dynamics** The two examples shown in Figs.2 and 3 exemplify the use of a SC for observing transient longitudinal dynamics. The first example shows a multibunch longitudinal instability in the Advanced Photon Source (APS) excited by a higher order mode in the accelerating cavities [9]. The oscillating pattern is created by longitudinal oscillation of bunches with relative phases of  $2\pi n_{bunch} f_{hom} / f_{rf}$ . The second example shows the decoherence of a longitudinal quadrupole oscillation from a mismatched beam at injection in the Advanced Light Source [10]. Immediately following injection, a quadrupole oscillation is evident. A few msec

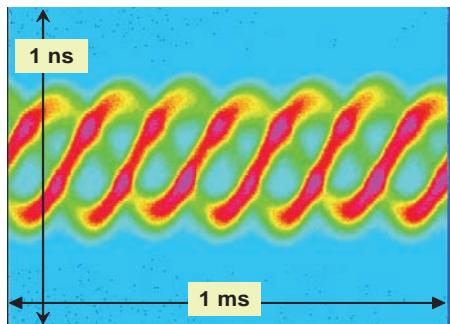


Figure 2: Streak camera image of a HOM-driven longitudinal multibunch instability [9]. The image shows individual bunches oscillating with a relative phase given by  $2\pi n_{bunch} f_{hom} / f_{rf}$ .

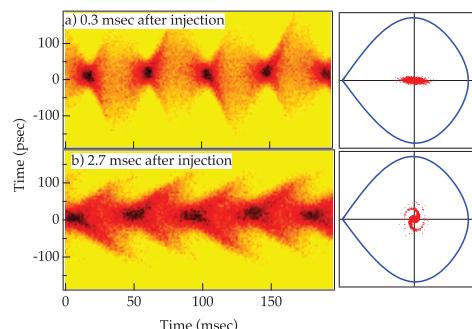


Figure 3: a) Measured SC image 0.3 ms following injection showing modulation of the bunch length. b) SC image 2.7 ms after showing filamentation of the large amplitude particles. [10]

later, the oscillation has begun to decohere from filamentation.

**Fast and slow head-tail motion** The examples shown in Figs.4 and 5 illustrate the use of a SC for observing time dependent transverse motion with fast and slow time scales. Figures 4a and b show the vertical motion along the length of a bunch following a vertical impulse kick to the beam in the presence of vertical chromaticity [11] measured at the APS. In the first five turns following the kick, there is no head-tail motion along the bunch. Half a synchrotron period later, significant head-tail motion has developed. In the second example, shown in Fig.5, the SC is configured with a relatively slow vertical sweep and even slower horizontal sweep to operate the SC

as a gated camera [12]. In this case, the SC captured a transient vertical instability along a bunch train driven by ions in at experiment at the Pohang Light Source.

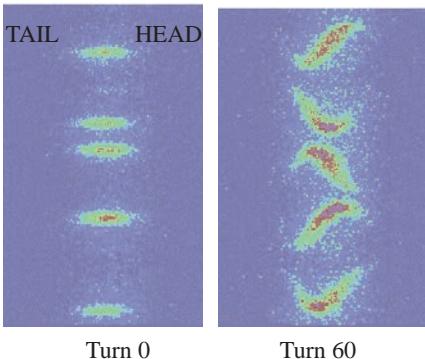


Figure 4: SC image of a transversely kicked beam showing the development of vertical head tail oscillations via chromaticity. The vertical axis is the vertical beam dimension and the horizontal axis is the fast time axis and shows the vertical beam position along the length [11]. Both a) and b) show the images from five successive beam passages.

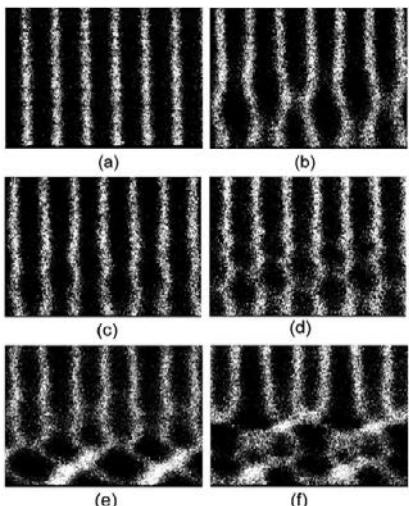


Figure 5: SC operated as a fast gated camera to observe the transient behavior along a bunch train for the fast beam ion instability at the Pohang Light Source. Measurement details provided in ref.[12].

## References

- [1] A.H. Lumpkin, BIW 90, AIP Conf. Proc. 229, 151 (1991)
- [2] E. Rossa, BIW 94, AIP Conf. Proc. 333, 11 (1994)
- [3] K. Scheidt, EPAC 2000
- [4] www.hamamatsu.com
- [5] hamamatsu.com; sydorinstruments.com; optronis.com; bifocompany.com
- [6] B.-L. Quian and H. E. Elsayed-Ali, JAP 91(1), 462-468 (2002)
- [7] B. J. Siwick et al JAP 92(3), 1643-1648 (2002)
- [8] G.A. Naylor et al, 2001 Meas. Sci. Technol. 12 1858
- [9] B. Yan, BIW 2006
- [10] J. M. Byrd, S. De Santis, PRST AB 4, 024401 (2001)
- [11] W. Guo et al, PRST AB 10, 020701 (2007)
- [12] J.Y. Huang et al, PRL 81, 4388 (1998)

### 7.4.12 Laser Wire

*M. Ross FNAL*

Laser based profile monitors, ('laser wires'), are used to reconstruct  $e^\pm$  and ion beam sizes using sampling, multi-bunch or multi-pulse, pass techniques [1, 2]. The beam transverse density is sampled through collisions with a focused laser pulse that has dimensions smaller than the beam. The relative scattering amplitude provides a measure of the beam. Laser wires are similar to scanning fiber or wire systems in which the fiber is moved through the beam (or vice versa) and the resultant scattering is detected in an associated detector. After data from a sequence of pulses (or bunches) is collected, the beam size may be estimated. Laser wire systems are an important beam size and emittance diagnostic for high energy linear colliders [3] and high power proton accelerators, such as the Spallation Neutron Source [4]. Laser wire systems have also proved useful in low emittance storage rings [5, 6].

The laser wire is a simple example of a laser-based profile monitor. In general there are 4 system components; (i) a source laser, (ii) a transport system to bring the laser beam to (iii) an optical demagnification system and (iv) a detector for Compton scattered photons and/or beam particles [7].

For given photon and electron beam characteristics the number of scattered particles may be estimated using the Compton scattering

Table 1: Example laser wire laser and particle beam parameters.

|                      | SLC IP [2] | SNS Linac [4] | ATF ring [5] | Petra II Ring [6] |
|----------------------|------------|---------------|--------------|-------------------|
| laser wavelength     | 350 nm     | 1064 nm       | 532 nm       | 532 nm            |
| laser pulse duration | 1e-10 s    | 1e-8 s        | cw           | 5e-9 s            |
| laser pulse strength | 0.001 J    | 0.05 J        | 100 W        | 0.06 J            |
| beam energy          | 45 GeV     | 0.2-1 GeV     | 1.3 GeV      | 7 GeV             |
| beam pulse duration  | 2e-12 s    | 3e-11 s       | 3e-11 s      | 4e-11 s           |
| beam size            | 600 nm     | 2 mm          | 1 micron     | 50 micron         |

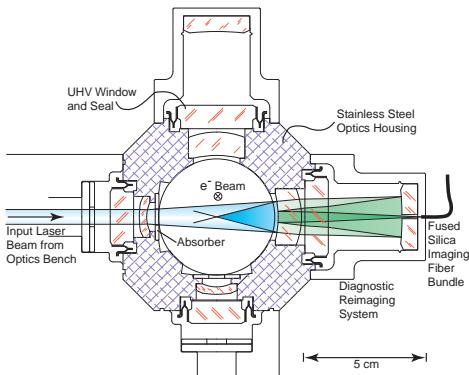


Figure 1: Laserwire IP. The particle beams pass into or out of the page and the laser light is introduced into the IP from the left side or from underneath for vertical or horizontal profile scans of the beam, respectively.

formula. For ion beams, such as H-, the scattering rate is given by the photo-detachment or photo-neutralization cross section. The beam size measurement resolution of the laser wire depends on the relative size of the photon and electron beam at the interaction point. Fig.1 shows a typical laser wire interaction region (IP) where the laser pulse collides with the particle beam at 90°.

Whereas the fiber in a wire scanner could be severed or otherwise damaged by the particle beam, the laser beam is not subject to such failure. This is a primary motivation for using a laser wire in a superconducting linac where fractured fiber segments could migrate to nearby superconducting RF cavities and reduce their performance. In addition, the density of the photon beam can be adjusted up to provide either a robust target capable of withstanding a very high power density beam in a high power e+/e- or ion linac or, alternatively, adjusted down to provide a thin scattering target that has no effect on the lifetime of a beam in a storage ring. Also, the minimum mode size can be made substantially smaller than the small-

est readily available fiber extending the resolution of the monitor well below other techniques. Typical minimum fiber sizes are a few microns in diameter, while laser beam minimum spot sizes ( $\sigma_0$ ) can be roughly ten times smaller, approaching the wavelength of light, with

$$\sigma_0 = \frac{\lambda f}{4\pi\sigma_{in}},$$

where  $\lambda$  is the wavelength of light,  $f$  is the focal length of the focusing system and  $\sigma_{in}$  is the input beam size. In the example of Figure 1,  $f = 30$  mm,  $\sigma_{in} = 2.5$  mm and  $\lambda = 350$  nm, giving  $\sigma_0 \approx 330$  nm. For the SLC IP laser wire, 350 nm light was chosen because it is available at high peak power, can be handled with high quality high power optical components and extends the endpoint of the Compton energy spectrum beyond that obtained from visible light.

Critical design and performance issues are (i) generation and control of the required photon density, (ii) preservation of the phase space of the laser beam through the transport system to the focusing interaction point optics and a system to verify that the phase space has been preserved, (iii) design, test and deployment of stable, aberration free focusing optics, and (iv) the placement and design of low background scattered beam detectors. Example laser and particle beam parameters are given in Table 1.

In addition, other practical considerations are: (i) the mechanism to rapidly move either the laser [8] or particle beam, (ii) phase-stable high pulse repetition rate and high power laser systems that are simple to operate and (iii) the usage of laser wires in conjunction with complementary diagnostics such as wire scanners, synchrotron light monitors or optical transition radiation screens that might be useful to optimise the laser wire Compton signal.

## References

- [1] T. Shintake, NIM A311 (1992) 453

## Sec.7.4: BEAM INSTRUMENTATION AND DIAGNOSTICS

- [2] R. Alley et al, NIM A379 (1996) 363
- [3] I. Agapov et al, PRST AB 10 (2007) 112801
- [4] Y. Liu et al, NIM A612 (2010) 241
- [5] Y. Honda et al, NIM A538 (2005) 100
- [6] A. Bosco et al, NIM A592 (2008) 162
- [7] A. Aryshev et al, EPAC08-TUPC011 (2008)
- [8] A. Bosco et al, APL 94 (2009) 1

### 7.4.13 Laser Interference Methods

*T. Shintake, SPring 8*

#### Transverse spot size measurement

This method has been used to measure a 70 nm beam at FFTB, SLAC.

Principle of operation Fig.1 shows the setup for spot size measurement schematically [1]. The laser beam is split into two by a half mirror, then focused into common foci, creating an interference fringe. We inject the electron beam on the focused laser beam from the normal direction. We measure the high energy  $\gamma$ -ray downstream, after the bending magnet that sweeps out the electrons to the beam dump. We slowly scan the electron beam by means of a weak steering magnet upstream, and measure the modulation depth in the  $\gamma$ -ray flux to determine the spot size.

The number of the  $\gamma$ -rays radiated by an electron in the laser beam  $\propto$  (transverse acceleration in the interference fringe field)<sup>2</sup>,

$$N_\gamma \propto B_0^2 (1 + \cos \theta \cos 2k_y y) \quad (1)$$

where  $\theta$  is the crossing angle of two laser beams and  $k_y$  is the wave number of the laser. The electron beam has a Gaussian distribution, which dilutes the modulation amplitude in the  $\gamma$ -ray. The modulation depth, defined by the peak-to-average divided by the average, is

$$M_0 = |\cos \theta| \exp \left[ -\frac{1}{2} (2k_y \sigma_y^*)^2 \right] \quad (2)$$

The fringe pitch is obtained from the wave number,

$$2k_y p = 2\pi, \quad p = \lambda_0 / 2 \sin(\theta/2) \quad (3)$$

For example, if we use a 1064-nm wavelength YAG-laser, and choose the crossing angle of 6°, we have 10.2- $\mu$ m pitch spacing. The finest pitch is 532 nm at 180° crossing angle. The smallest spot size that can be detected by the finest fringe is  $\sigma_{y,\min} = 0.04\lambda_0$  which corresponds to 90% modulation.

The available number of the  $\gamma$ -ray and the energy spectrum can be estimated by the Compton scattering cross-section with relativistic correction [2].

Correction factors By taking into account two major error sources: the power imbalance of two laser beams and laser beam thickness along the electron trajectory, the modulation depth becomes

$$M = M_0 \frac{2\sqrt{P_2/P_1}}{1 + P_2/P_1} \frac{1}{\sqrt{1 + (2k_y \sigma_y^*)^2 (\sigma_L/\beta_y^*)^2}} \quad (4)$$

where  $P_2/P_1$  is the power ratio of laser beams,  $\sigma_L$  is the laser beam width,  $\beta_y^*$  is the beta-function of the electron beam.

Measurement example Fig.2a shows measured data in the FFTB beam line at SLAC [3, 4]. The solid line is the least-squares fit. The modulation depth is 0.68, and the spot size is  $\sigma_y^* = 66$  nm. Fig.2b shows the spot size distribution for a 3 hour measurement. As the laser beam thickness is neglected in this plot, the actual size is 6 to 7 nm smaller than shown.

#### Features of the interferometric size monitor

(i) The measurement mechanism is simple and easy to understand; (ii) High power beam capability. The wire scanner using carbon fibers is subject to being broken by a high power electron beam, the laser beam is not; (iii) Well defined measurement scale. The fringe pitch is defined by the wavelength of the laser beam and the crossing angle; (iv) The interference fringe is quite stable; (v) The reading errors are always on the safe side. Every source of measurement error produces positive errors on the measured spot size; (vi) Wide range of measurement. Changing the crossing angle, we can choose a fringe pitch in over a range from mm to nm.

**Bunch length measurement** This is proposed as a complementary method to the spot size monitor. (See also Sec.)

Principle of operation Fig.3 shows the schematic diagram of the bunch length measurement system [4]. By mixing two laser beams of different frequency they will beat to create an intensity modulation with a frequency equal to the frequency difference between the two laser beams. We inject electron bunches into the laser beam from the normal direction, and measure the Compton scattered  $\gamma$ -rays downstream. Since the amplitude of the electric field in the beat-wave is periodically modulated, the  $\gamma$ -ray flux will change according to the injection phase of the electron bunch.

Bunch spectrum measurement The combined laser power is amplitude modulated with the

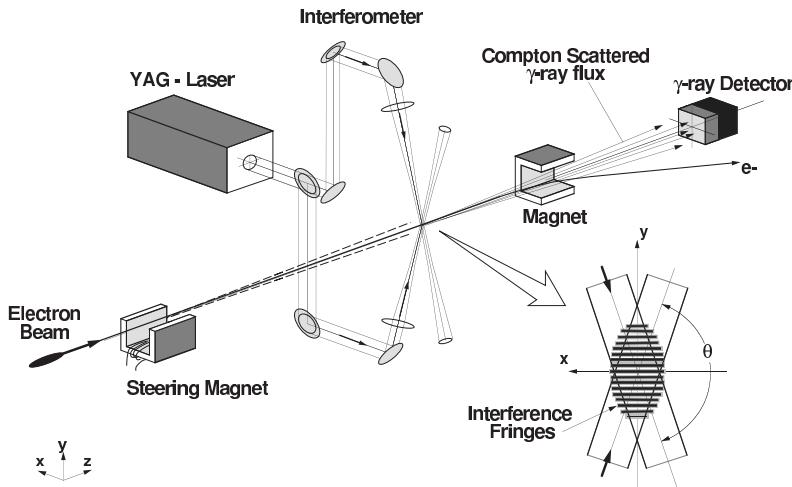


Figure 1: System diagram of the spot size monitor.

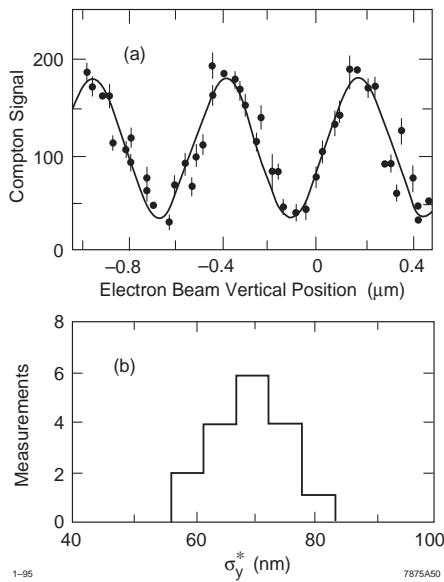


Figure 2: Example measurement and analysis.

difference frequency

$$P(y,t) = P_1 + P_2 + 2\sqrt{P_1 P_2} \cos 2\pi f_b(t + \frac{y}{c}) \quad (5)$$

where  $f_b = f_2 - f_1 = f_0 (\Delta\lambda/\lambda)$ ,  $\Delta\lambda = \lambda_1 - \lambda_2$  is the wavelength difference between two lasers. Assuming a Gaussian electron bunch, and Gaussian laser beam, the variation of the  $\gamma$ -ray flux

becomes

$$N_\gamma \propto N_e P_0 \left[ F_0 + e^{-\frac{(k_b \sigma_\Sigma)^2}{2}} F_\omega \sin(\phi_b + \phi_f) \right] \quad (6)$$

where  $\sigma_\Sigma^2 = \sigma_y^2 + \sigma_{Lz}^2$ .  $\sigma_y$  is the beam height and  $\sigma_{Lz}$  is the laser beam width.  $F_0$  and  $F_\omega$  are the Fourier components of the electron bunch. The modulation depth is

$$M \equiv \frac{N_{\gamma \max} - N_{\gamma \min}}{N_{\gamma \max} + N_{\gamma \min}} = \frac{F_\omega}{F_0} e^{-\frac{(k_b \sigma_\Sigma)^2}{2}} \quad (7)$$

From the modulation depth, the bunch shape is determined.

Practical parameter example Tab.1 displays parameters of interest for a measurement of the bunch length at the exit of a bunch compressor for a linear collider.

Table 1 Practical parameters

|                |                 |              |
|----------------|-----------------|--------------|
| Electron Beam  |                 |              |
| Energy         | E               | 1.75 GeV     |
| Size           | $\sigma_x$      | 118 μm       |
|                | $\sigma_y$      | 19 μm        |
|                | $\sigma_z$      | 85 μm        |
| Laser Beam     |                 |              |
| Wavelength     | $\lambda$       | 532 nm       |
| Tuning range   | $\Delta\lambda$ | 0 - 1 nm     |
| Beat Frequency | $f_b$           | 0 - 1 THz    |
| Width          | $\sigma_{Lz}$   | 30 μm        |
|                | $\sigma_{Lx}$   | 200 μm       |
| Power          | $W_l$           | 100 mJ/10 ns |

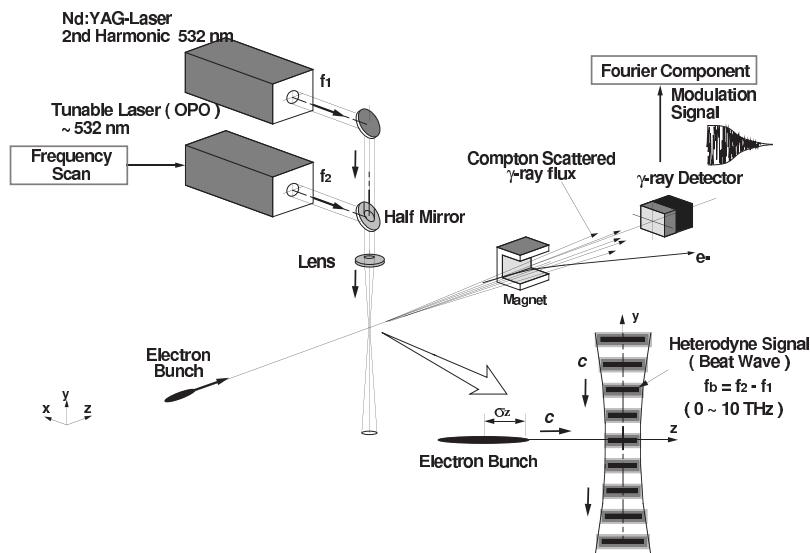


Figure 3: Schematic diagram of bunch length measurement system using the laser heterodyne.

Accuracy There are two issues for accurate measurement of the bunch spectrum, i.e. the scale accuracy and the amplitude accuracy. The longitudinal scale accuracy in bunch length measurement is determined by the frequency accuracy of the tunable laser. For the tunable laser source, we can use OPO: Optical Parametric Oscillator [5, 6]. Using an injection seeded technique, the line width can be lowered to only 0.006 nm at 532 nm, which provides the beat signal of 500 GHz with 1% frequency accuracy. It corresponds to the longitudinal scale accuracy of 1%. The amplitude accuracy is determined by the statistical error in the  $\gamma$ -ray measurement, which depends on the total flux of the  $\gamma$ -ray per bunch. In the present example, the average number per bunch is expected to be a few thousand, thus the statistical error can be < a few %. The laser power and the electron charge fluctuations can be corrected by normalization with monitoring the power and the charge.

Features of the method (i) Non-destructive monitor. No foil nor wire is needed. The laser beam does not destroy the electron beam; (ii) The laser beam performance (intensity, spot shape, and beat frequency, etc.) is independent of the electron beam. They are well isolated; (iii) Bunch selection capability. The YAG laser and OPO laser can generate a short laser-pulse at a few ns,

which enables one to pick-up one bunch from the multi-bunch train; (iv) Accurate time-base at THz frequency. The beat-wave frequency is defined by the laser wavelength, which is well controlled in a modern OPO lasers and can be calibrated by a grating spectrometer.

## References

- [1] T. Shintake, NIM A311 (1992) 453
- [2] V.I. Telnov, Int. Workshop on the Next Generation of Linear Colliders, SLAC (1988)
- [3] B. Schwarzschild, Physics Today, July 1994, p.22
- [4] T. Shintake, Proc. 7th BIW, AIP Proc. 390 (1996) p.130
- [5] S.E. Harris, Proc. IEEE, Vol.57, No.12 (1969) p.2096
- [6] Spectra-Physic Lasers, 1330 Terra Bella Avenue, Mountain View, CA, 94039-7013 USA

### 7.4.14 Beam Size via Synchrotron Radiation *T. Shintake, Spring 8*

**Gradient Undulator** Fig.1 shows an idea to measure the transverse beam size using x-ray radiation from the gradient undulator [1]. The gradient undulator has a special configuration, i.e. the longitudinal position of the periodic magnets are in the same phase at both poles. Therefore, there is no transverse field on the axis due to field

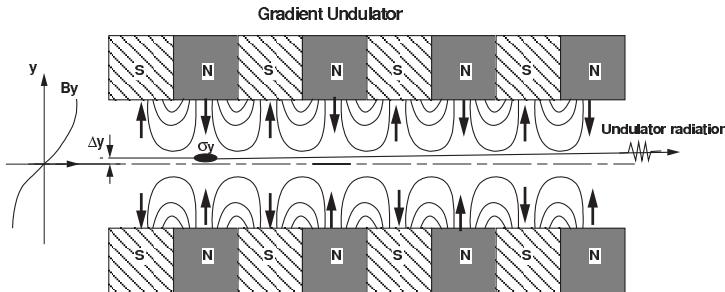


Figure 1: Beam size measurement using gradient undulator.

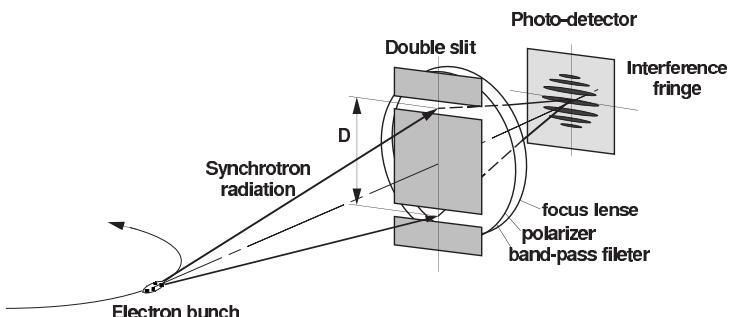


Figure 2: Beam size measurement by interference fringe of SR.

cancellation. The transverse magnetic field  $B_y$  is proportional to the offset from the axis and the radiation power from a charged particle running in the gradient undulator  $\propto (\text{offset } \Delta_y)^2$ . In the case of an actual beam, the transverse size is finite, thus the total number of the radiation per bunch becomes

$$N_\gamma = aN(\sigma_y^2 + \Delta_y^2) \quad (1)$$

where  $a$  is a constant and  $N$  is the number of particles. By scanning the vertical position of the undulator using a mover, we can calibrate Eq.(1) to determine the constant  $a$ . We can obtain the spot size from the minimum value of  $N_\gamma$ . This method is considered to be used in the main linac of an  $e^+e^-$  linear collider project, where the typical rms beam-size is  $8 \mu\text{m}$ . The parameter of the gradient undulator is 1 m long, 1 cm period and 3 T/cm gradient. The most technical challenge is to develop such undulator at high precision. To measure  $8 \mu\text{m}$  beam with 3% accuracy, the straightness of the magnetic center has to be better than  $2 \mu\text{m}$  for 1 m long undulator.

**Double-Slit Interferometer** Mitsuhashi applied the double-slit interferometer to measure the transverse size of electron beam in a storage ring [2]. Fig.2 shows his experimental set up, schematically. This method was originally developed by Michelson to measure the angular diameter of stars in 1920, which is well known as Michelson's stellar interferometer [3].

The electron beam running on a curved trajectory in a storage ring radiates the SR, which splits in two beams by the double slit, and is then to a common focus by a lens. A band-pass filter selects a quasi-monochromatic light from the wide-band SR, thus the two beams cause interference and produce a fringe pattern on the photo-detector.

Since, each electron in the bunch emits the SR without correlation, creating a fringe pattern at different phase, by overlapping SR from many electrons, the contrast of the fringe will fade-out. For a small beam size, the contrast will be high, but for a large beam the contrast will be low. Therefore, from the fringe contrast, we can

determine the transverse size of the electron beam. The fundamental theory of this measurement is known as the Van Citter-Zernike's theorem [4].

The resolution of this method is approximately estimated by

$$\delta_y \approx [L_1/D] \lambda \quad (2)$$

where  $L_1$  is the distance from the electron beam to the slit,  $D$  is the double-slit spacing and  $\lambda$  is the wavelength. In practice,  $L_1$  is typically 5 to 10 m.  $D$  is limited by the aperture of the light guide and the focal lens, for example it is about 0.1 m. If we use visible light at 500 nm, the resolution becomes 2.5  $\mu\text{m}$ , and ultraviolet light will improve it down to 1  $\mu\text{m}$ .

Features of this method are (i) the apparatus is not expensive, (ii) wide-range beam size measurement is possible, and (iii) it is a non destructive measurement.

Recently an improvement in resolution was made using vertically polarized radiation as described in A. Anderson et al, NIM A 591 (2008) 437.

## References

- [1] E.G. Bessonov et al, Internal DESY report M96-18 (1996)
- [2] T. Mitsuhashi, PAC 97
- [3] M. Mansuripur, Optics & Photonics News, OSA: Optical Society of America, Vol.9, No.1 (1998)
- [4] M. Born, E. Wolf, Principle of Optics, Ch.X, Pergamon Press (1980)

### 7.4.15 Electron Cloud Measurements

*F. Caspers, F. Zimmerman CERN*

Electron cloud build-up mechanisms are discussed in Sec.2.4.14. Here we focus on measurements of electron cloud for proton and positron (or electron) beams.

**Pressure rise** Electron cloud build up often manifests itself through a significant increase of the vacuum pressure by several orders of magnitude above a certain threshold beam current [1, 2, 3]. Those pressure rise signatures can be observed by monitoring vacuum gauges, vacuum pump currents or residual gas analyzers. The typical time constant of pressure rise detectors is between milliseconds and seconds and permits a good correlation between the formation of a e-cloud e.g., during the magnetic cycle of a ramped

machine or when injecting several bunch trains with a time interval of seconds. Pressure-rise diagnostics, which could be selective for certain residual-gas components may also be applied in stand-alone test units to provide evidence for the onset of multipactoring.

**Beam instabilities, transverse feedback & head-tail monitor** The electron-cloud wake field drives single- [2, 4, 5] and coupled-bunch instabilities [4, 5, 6, 7]. Transient oscillations, bursting beam-size blow up as well as bunch-dependent synchro-betatron sidebands have been identified as signatures of electron-cloud driven single-bunch instability [8]. In proton rings, the single-bunch instabilities can also directly be observed using a head-tail monitor [4, 5] and might be suppressed by increased chromaticity [5], Landau-damping octupoles, and transverse feedback [4, 5, 9, 10]. In positron storage rings, the electron cloud density may be reduced, and thereby instabilities be avoided, by installation of weak solenoids in field-free regions [11], the effect of which is another indicator of electron cloud. The electron-cloud driven coupled-bunch instabilities can be monitored and suppressed by a transverse bunch-by-bunch feedback system [5, 8].

### Beam size increase, synchrotron-light monitor

In positron storage rings, where coupled bunch-instabilities are efficiently suppressed by feedback, the electron cloud can manifest itself in a vertical beam-size increase with current above a certain threshold current that varies with the bunch spacing [10]. Electron cloud related increase in vertical beam-size has been observed in KEKB [12] via synchrotron light interferometers as function of beam intensity and with e-cloud suppressing solenoids in arcs and straight sections being turned on and off, respectively. As a result, it could be shown that the solenoids in the straight sections did contribute to the mitigation of the electron-cloud impact.

**Beam lifetime, beam losses, bunch length** The nonlinear pinch of the electron cloud during a bunch passage leads to highly nonlinear fields and can be responsible for incoherent particle losses, bunch-length shrinkage, and a poor beam lifetime [13, 14]. The lifetime can be monitored by fast beam-current transformers and the bunch length by a fast wall-current monitor or a streak camera. Transverse Schottky monitors also provide

further insight in the incoherent effect of the electron cloud.

**Tune shift** Accumulation of electrons leads to tune shift along the bunch train, the magnitude of which indicates the integrated strength of the electron cloud around the storage ring [4, 5, 10, 8, 15, 16]. An example of this tune shift vs time in the presence of an electron cloud is shown in Fig.1 [17].

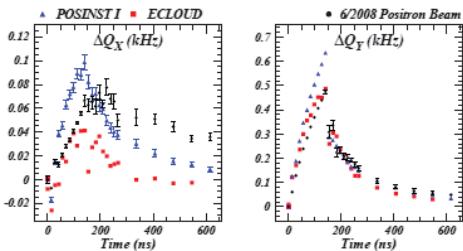


Figure 1: Comparison of measured and modeled tune shift along a bunch train in CesarTA for 5.3 GeV positrons [17]. The bunch train was about 140 ns long (10 bunches spaced by 14 ns). The tune shift behind the bunch train was measured with the help of a witness bunch placed at a variable distance.

**Retarding field analyzer (RFA) and shielded button pick-up** The RFA is one of the well known dedicated diagnostic elements [18]. An RFA basically consists of a triode like structure. A control grid is situated between a shielding grid at ground potential which is flush mounted to the beam-pipe inner surface. The control grid can be set to a variable potential. Depending on its voltage setting, the energy selection of electrons arriving at the anode (=collector) takes place. Typical values for this control grid setting are between 0 and 1 kV. The energy spectrum of the electrons can be deduced from the derivative of the collected current vs time for a swept (vs time) control grid voltage. The same structure, just without a control grid, i.e. electrically biased button electrode covered by a grid at ground potential, is referred to as a “shielded pick-up” [19].

The electromagnetic shielding by some set of fine wire meshes is mandatory since otherwise beam induced signals would perturb the response. In Fig.2 a shielded button pick-up (left) as well as an RFA are depicted. An example for bunch to bunch electron cloud build-up during the batch and the successive decay is depicted in Fig.3.

**Calorimetry** Heat deposited by the cloud electrons on the inner beam-pipe wall, most relevant for cryogenic environments (SC magnets), can be measured in dedicated set ups at room temperature (WAMPAC) [20] or at cryogenic temperature (COLDEX) [21], as well as by smaller pick-up calorimeters [22]. An example of a pick-up calorimeter is shown in Fig.4. The actual calorimeter has an internal temperature reference, and internal resistor for in-situ thermal calibration and a platinum resistor for the measurements. Due to the small mass of the active elements the time constant is around 3 minutes. Its very high sensitivity permits extrapolation of measured data to heat loads below 0.1 Watt/meter on the vacuum chamber. As an alternative, cryogenic heat load measurements have been tested using an actively cooled beam-screen of about 1 meter length operating in the temperature range of 5 to 100 K [21].

**In situ secondary emission yield curve (SEY) measurement** The secondary emission yield curve and its time evolution (“scrubbing effect”) can be measured in situ [1]. For such measurements a probe electron beam with adjustable energy in the range from a few eV to about 10 keV is impinging on a surface to be measured and the scattered electrons are collected by a suitably biased electrode. The latter is comparable to a conventional ex-situ test stand in the lab. The sample under test, which in beam operation forms part of the beam-pipe surface, is mounted on a drum. The drum can slowly be rotated through some remotely controlled mechanical actuator so that it faces the external electron gun for dedicated SEY measurement. The change in SEY is measured in suitable time steps after exposure to the electron cloud of the circulating proton beam (Fig.5). This measurement is performed over a certain time period (days).

**Multi-strip detectors** Multi-strip detectors can be considered as an advanced version of shielded buttons [1, 22]. The bottom of a flat beam pipe contains a section with a grid or just many small holes. Behind this grid or holes (perforated surface) not a single collector is present but instead a large number (e.g. 50) of parallel metallic strips with individual readout and either common or individual bias (Fig.5). With this configuration 2D distributions of the electron cloud density and their time evolution have been measured, in field-free regions, in bending magnets and in quadrupoles [23]. Various spatial structures have

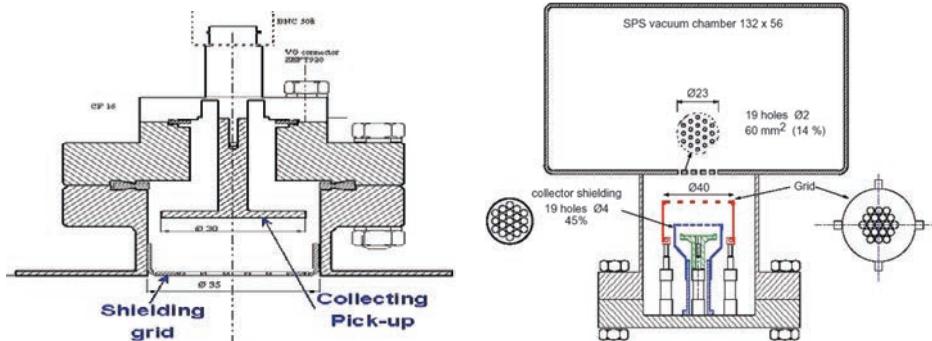


Figure 2: A shielded button pick-up (a) and an RFA (b).

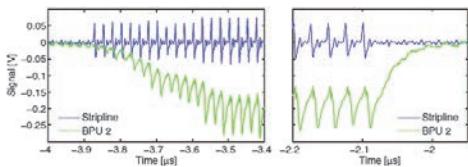


Figure 3: Bunch to Bunch electron cloud buildup (left) and decay (right) measured with a button pickup during the second last turn of the LHC beam in the CERN-PS. The 25 ns gap between successive bunches can be clearly resolved on the strip-line electrode signal.

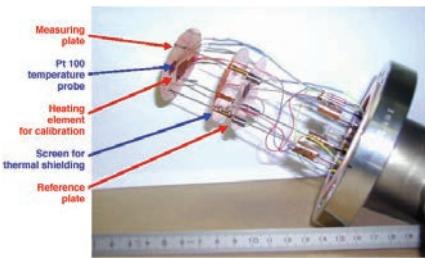


Figure 4: Pickup calorimeter deployed in the CERN SPS [22].

been observed (2, 3, or 4 stripes in dipoles, and 4 or 8 stripes in quadrupoles) [24].

**Microwave transmission** This method integrates the electron cloud density over the length of one or several bending magnets and drift spaces [25, 26, 27]. Very low RF power (typically a fraction of a Watt) is coupled via suitable antennae to a waveguide so as to propagate along the beam-pipe (Fig.6). Usually TE-waveguide modes are chosen in order to minimize direct coupling with

the relativistic particle beam and since TE-modes normally are the fundamental modes in a beam-pipe. This can reduce complications with mode selective excitation and pick-up structures which should be minimized in their direct coupling to the TEM like field of the relativistic beam. It should be underlined that this method is able to integrate over the length of several bending magnets and drift spaces. For typical electron cloud densities in the order of  $10^{12} m^{-3}$  the microwave frequency is well above the plasma frequency  $\omega_p$  and one can diagnose the electron density via a faint phase shift given by [28]

$$\Delta\varphi = -L\omega_p^2 / \left( 2c\sqrt{\omega_{rf}^2 - \omega_c^2} \right) \quad (1)$$

Here  $\omega_c$  stands for the cutoff frequency of the waveguide mode, and  $L$  is length of the beam pipe section. The plasma frequency  $\omega_p$  ( $m s^{-1}$ ) is related to the electron density by  $\omega_p = 56.6\sqrt{n_e}$ , where  $n_e$  is the electron density per cubic meter. This equation is only valid for a cold plasma near thermodynamical equilibrium and assuming a homogeneous e-cloud. As for absolute conversion of a measured phase shift into e-cloud density we must keep in mind that the e-cloud is non-uniform and its energy spread far from thermal equilibrium. Some refinements of the first order model (1) are discussed in [29]. In case a static magnetic field is present (e.g. a bending field) we obtain a signal enhancement factor in the vicinity of the cyclotron resonance (28 GHz/Tesla) provided that also an electric microwave field component transverse to the static magnetic field is present. The enhancement factor  $F_e$  is proportional to

$$F_e \propto \frac{1}{1 - \left( \frac{eB}{\omega_{rf}m_e} \right)^2} \quad (2)$$

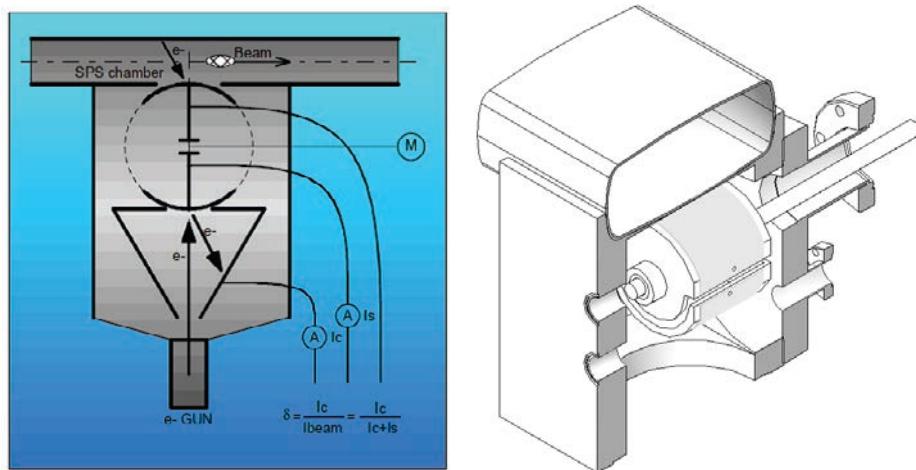


Figure 5: In situ SEY measurement at the SPS [1].

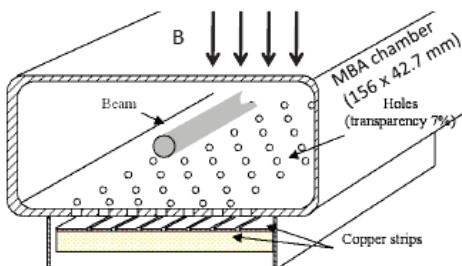


Figure 6: Schematic of multi-strip detector at the SPS [1].

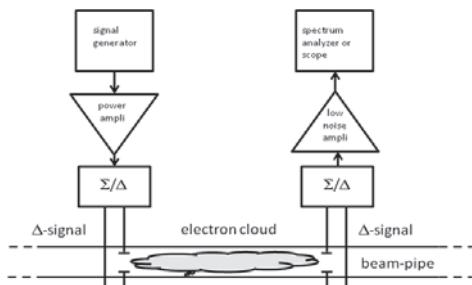


Figure 7: Schematic of microwave transmission measurement for electron-cloud diagnostics.

In many machines, however, it is not possible simultaneously to stay in the fundamental mode of propagation (low GHz range) and to meet the cyclotron resonance condition at a given, rather high bending field. Caution is advised to carefully control possible inter-modulation effects in the electronic signal transmission chain from the beam-induced signal. Also inter-modulation and/or hardware transfer function related amplitude modulation (AM) to phase modulation (PM) conversion may give rise to fake PM signals which would lead to an overestimation of the true results. For rejection of undesired AM components and for filtering out the PM of interest, the analog demodulation function available on certain vector spectrum analyzers can be used. But even this AM re-

jection function cannot eliminate the above mentioned fake PM signals.

Other caveats include: Microwave measurements have been plagued by uncertainties due to reflections of the microwave signal in the beam-pipe further up- and downstream thus leaving a question mark on the effective length of the measurement section. Also minor changes in cross-section of the beam-pipe can lead to considerable uncertainties in particular when operating close to waveguide cutoff. As for the actual measurement techniques one should mention two basic methods: Evaluation of the phase modulation (PM) sideband (with suppression of beam related AM) [27, 29] by means of a vector signal analyzer or direct phase vs time reading from a sensitive phase detector [30].

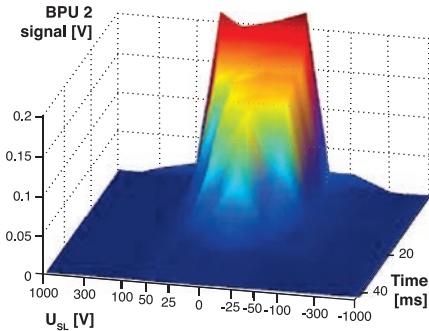


Figure 8: Electron cloud button pickup signal measured in the CERN PS (straight section 98) for the nominal LHC beam. Color coding shows the intensity of the electron cloud as a function of the applied voltage on the clearing electrode.

**Clearing electrodes** An unshielded stripline electrode at about 1 kV (positive or negative bias voltage) can clear the electron cloud [31, 32] (Fig.8). With a positive bias voltage it collects most of the primary electrons [33].

There are many ways to implement such clearing electrodes. Apart from the classical metallic strip-line structures which also have a considerable beam coupling impedance, there exist concepts for, with respect to impedance, nearly invisible clearing electrodes. These structures use a thin (50 to 500 micron) dielectric layer (alumina or enamel with some very thin (compared to the skin depth at the highest frequency of interest) resistive layer [34, 35]). Another approach has been used successfully at KEK [36].

## References

- [1] M. Jimenez et al, ECLOUD02 CERN Geneva, p.17, CERN-2002-001
- [2] F. Zimmermann, PRST AB 11, 041002 (2008)
- [3] W. Fischer et al, PRST AB 11, 041002 (2008)
- [4] K. Cornelis, ECLOUD02 CERN, p.11, CERN-2002-001
- [5] G. Arduini et al, PAC03 p.3038
- [6] M. Izawa et al, PRL 74:5044-5047, 1995
- [7] K. Ohmi, PRL 74:1526-1529, 1995
- [8] J.W. Flanagan et al, PRL 94:054801, 2005
- [9] J.D. Fox et al, IPAC10, p.2806
- [10] H. Fukuma, ECLOUD04 CERN-2005-001, p.15
- [11] M. Tobiayama et al, PRST AB 9:012801, 2006
- [12] H. Fukuma et al, EPAC 2000, WEP5A
- [13] E. Benedetto et al, PRL 97:034801, 2006
- [14] E. Shaposhnikova, CARE-HHH CERN-GSI meeting 20.06.2006
- [15] K. Ohmi et al, APAC 01 p.445
- [16] M.A. Palmer, ECL2, Geneva, CERN-AB-2007-064, 2007, p.74
- [17] J.A Crittenden et al, IPAC10, TUPD024
- [18] R. Rosenberg, K. Harkay, NIM A 453 (2000) p.507
- [19] J.M. Jimenez et al, PAC03 (2003) 307
- [20] V. Baglin and B. Jenninger, PRST AB 6, 063201, 2003
- [21] V. Baglin et al, LHC-Project-Report-667 (2003)
- [22] N. Hilleret, Mini-Workshop on SPS Scrubbing Run Results and Implications, CERN, 28 June 2002
- [23] G.F. Dugan et al, ICFA Beam Dynamics Newsletter No. 50, 2009, p.16
- [24] J.M. Jimenez, Proc. CARE-HHH Chavannes-de-Bogis, CERN-2009-004, p.120
- [25] T. Kroyer et al, ECLOUD04, CERN-2005-001, p.89
- [26] S. De Santis, PRL100:094801, 2008
- [27] F. Caspers, F. Zimmermann, PAC09, WE1PBI02
- [28] K. Sonnad et al, PAC 07 THPAS008
- [29] S. Federmann et al, IPAC10, TUPEA076
- [30] N. Eddy et al, PAC09, WE4GRC02
- [31] O. Bruning, LHC Project Report 158, 1997. Magnetic Field and Image ChargesMagnetic Field and Image Charges
- [32] F. Zimmermann, CERN SL-Note-2001-022-AP, 2001
- [33] E. Mahner et al, PRST AB 11, 094401 (2008)
- [34] F. Caspers et al, PAC2007, WEOAC02
- [35] Y. Suetsugu, IPAC2010, WEOAMH01
- [36] Y. Suetsugu et al, NIMA 598 (2009) 372
- [37] R.J. Macek et al, JAP 102:124904, 2007
- [38] S.de Santis et al, IPAC10, MOPE088
- [39] P. Lebrun et al, PAC09, TH5PFP019

### 7.4.16 Beam Loss Monitors

R. Jones, CERN  
K. Wittenburg, DESY

Beam loss monitors (BLMs) [1] have three main uses in particle accelerators:

**• Damage prevention** - Beam loss may result in damage to accelerator components or the experimental detectors. One task of any BLM system is to avoid such damage. In some accelerators it is an integral part of the protection system, signalling the beam abort system to fire if a certain loss rate is exceeded. This is of vital importance to the

new generation of superconducting accelerators, for which even fairly small beam losses in the superconducting components can lead to magnet quenches.

- **Diagnostics** - Another task of BLM systems is to identify the position and time of unacceptable beam losses and to keep the radiation level in the accelerator and its surroundings as low as possible.
- **Luminosity optimisation** - BLMs can also help in the tuning of the machine in order to produce the long lifetimes necessary for improved luminosity.

The task of a BLM system is to establish the number of lost particles at a certain position within a specified time interval. Most BLM systems are mounted outside the vacuum chamber, so that the detector normally observes the shower caused by the lost particles interacting in the vacuum chamber walls or in the materials of the magnets. The energy loss of incident charged particles scattering on atomic electrons is described by the Bethe-Bloch equation. The minimum in  $dE/dx$  is typically  $1\text{-}2 \text{ MeV.cm}^2.\text{g}^{-1}$ , usually occurring between 1 and 5 GeV for protons. Defining a Minimum Ionising Particle (MIP) as having an energy loss of  $2 \text{ MeV.cm}^2.\text{g}^{-1}$  allows correspondence between an energy deposition of 1 Gy in a BLM to a charged particle flux of  $3.1 \text{ MIPs.cm}^{-2}$ . The actual proportionality between the number of incident particles and the signal generated by the BLM depends on the position of the BLM with respect to the beam, the type of lost particles and the intervening material. It also depends on the momentum of the lost particles, which may vary by a large amount during the acceleration cycle.

#### 7.4.16.1 Beam loss monitoring using ionisation detection

When a charged particle passes through matter the medium will be ionised, producing ion-electron pairs. The amount of energy loss in creating an ion-electron pair is given by the ionisation constant. Similarly a charged particle traversing a semiconductor such as silicon will create electron-hole pairs. The number of pairs produced per MIP per cm is given by the following

equation:

$$\begin{aligned} \text{N per cm per MIP} &= \frac{1}{\text{Ionisation Constant (eV)}} \\ &\times 2 \cdot 10^6 \left( \frac{\text{eV.cm}^2}{\text{MIP.g}} \right) \\ &\times \text{Density} \left( \frac{\text{g}}{\text{cm}^3} \right) \end{aligned}$$

For a gas such as Nitrogen at atmospheric pressure this yields  $\sim 70$  ion pairs per cm per MIP, while for silicon this yields  $\sim 10^6$  electron-hole pairs per cm per MIP, showing the much larger charge production obtained in solid-state ionisation chambers. The sensitivity of a given medium is given by:

$$\begin{aligned} \text{Sensitivity} \left( \frac{\text{C}}{\text{Gy.cm}^3} \right) &= 10^{-3} \left( \frac{\text{J}}{\text{g}} \right) \\ &\times \text{Density} \left( \frac{\text{g}}{\text{cm}^3} \right) \\ &\times \frac{1}{1.6 \cdot 10^{-19}} \left( \frac{\text{eV}}{\text{J}} \right) \\ &\times \frac{1.6 \cdot 10^{-19}(\text{C})}{\text{Ionisation Constant (eV)}} \end{aligned}$$

**Long ionisation chambers** In 1963, Panofsky [2] proposed a BLM system for SLAC which consisted of one long (3.5 km) hollow coaxial cable filled with Ar (95%) + CO<sub>2</sub> (5%), mounted on the ceiling along the LINAC, about 2m from the beam. When a beam loss occurs, an electrical signal is produced which propagates to both ends of the cable. Position sensitivity is achieved by comparing the time delay between the direct pulse from one end and the reflected pulse from the other. The time resolution is about 30 ns ( $\sim 8$  m) which, for shorter versions, can be reduced to about 5ns. This principle of space resolution works for linear accelerators and transport lines with a bunch train much shorter than the machine and with relativistic particles. For particles travelling significantly slower than the signal in the cable the resolution of multiple hits in the cable becomes difficult. In this case, and for circular machines, it is necessary to split the cable. Each segment has to be read out separately, with a spatial resolution which becomes approximately equal to their length.

**Short ionisation chambers** Short ionisation chambers are used in many accelerators [e.g. [3, 4]]. They are more or less equally spaced along the accelerator, typically near the quadrupoles

| BLM Type                              | Acquisition Type        | Typical Sensitivity ( $\text{nC.Gy}^{-1}$ ) | Typical Dynamic Range | Time Resolution        | Spatial Resolution | Typical Radiation Tolerance | Typical Size        |
|---------------------------------------|-------------------------|---|-----------------------|------------------------|--------------------|-----------------------------|---------------------|
| Short Gas Ionization Chamber          | Integrating             | $10\text{-}10^2 \text{ cm}^{-3}$            | $< 10^8$              | $\mu\text{s}$          | Detector location  | High                        | $< 1 \text{ m}$     |
| Long Gas Ionisation Chamber           | Integrating             | $10\text{-}10^2 \text{ cm}^{-3}$            | $10^4$                | $\mu\text{s}$          | $< 10 \text{ m}$   | Medium                      | Few m to several km |
| Solid State Ionisation Chamber        | Counting                | $10^4 \text{ cm}^{-2}$                      | $10^9$                | ms                     | Detector location  | Medium                      | $< 10 \text{ cm}$   |
| Plastic Scintillation Detectors       | Integrating or Counting | $10^3\text{-}10^6 \text{ cm}^{-3}$          | $< 10^8$              | ns (integ.) ms (count) | Detector Location  | Low                         | $< 30 \text{ cm}$   |
| Inorganic Scintillation Detectors     | Integrating or Counting | $10^4\text{-}10^8 \text{ cm}^{-3}$          | $< 10^8$              | ns (integ.) ms (count) | Detector Location  | Medium                      | $< 30 \text{ cm}$   |
| Aluminium Cathode Electron Multiplier | Integrating             | $10^5 \text{ cm}^{-2}$                      | $10^5$                | ns                     | Detector location  | Medium                      | $< 10 \text{ cm}$   |
| Secondary Emission Monitor            | Integrating             | $10^{-3} \text{ cm}^{-2}$                   | $10^5$                | ns                     | Detector location  | High                        | $< 50 \text{ cm}$   |
| Optical Fibre Detector                | Integrating or Counting | $10^5 \text{ cm}^{-2}$                      | $10^4$                | ns (integ.) ms (count) | $< 1 \text{ m}$    | Medium to High              | Few m to several km |

where the beta function is a maximum, with additional units at special positions such as aperture restrictions, targets, collimators, etc. The chamber provides some medium with which the secondary particles created by the beam loss can interact, typically filled with a gas such as nitrogen or argon. This interaction produces electron-ion pairs which are collected by one or more high voltage gaps along the length of the chamber. The resulting current is then measured and is proportional to the beam loss at the location of the monitor. Secondary Electron Emission Monitors (SEM), consisting of one or more signal electrodes under vacuum, are employed when the expected loss rates are very high [5]. SEM typically have a  $10^5$  times lower sensitivity than standard ionization chambers, but a much faster response time.

**PIN photodiodes** For circular electron accelerators which emit hard synchrotron radiation it is difficult to distinguish between the beam

loss distributions and the synchrotron radiation background using traditional BLM techniques. In DESY-HERA, an electron-proton collider, the warm electron and superconducting proton rings were in the same tunnel. Protection of the superconducting proton beam magnets from beam loss induced quenches therefore relied on a BLM system which sees only the proton beam losses and not the synchrotron radiation background. In this case back to back PIN photodiodes were used to distinguish between the hadronic shower created by beam losses and the synchrotron radiation [6]. The charged particles will interact with both photodiodes, giving a coincidence signal, while a photon will be absorbed by one diode only. In contrast to the charge detection of most other BLM systems, PIN photodiode detection depends on counting coincidences, with the count rate proportional to the loss rate so long as the number of overlapping coincidences is small.

**Aluminium cathode electron multipliers** In such detectors the sensitivity of photomultipliers to ionising radiation is increased by replacing the photocathode with an aluminium foil. This foil then works as a secondary electron emitter when irradiated. A BLM system consisting of Aluminium Cathode Electron Multipliers (ACEMs) is installed on the CERN PS and PS-Booster [7] and on FLASH at DESY [8]. It is very fast, with signal rise times in the order of 10 ns, but is rather expensive since the ACEM is not a standard tube of photomultiplier manufacturers.

#### 7.4.16.2 Beam loss monitoring using light detection

**Scintillation counters** Beam losses can also be measured through the detection of particle induced light production in scintillators or Cherenkov radiators. Scintillation systems are very sensitive, but are often prone to radiation damage and suffer from large unit to unit gain variation. Liquid and crystal scintillators have, nevertheless, been employed as beam loss systems in various accelerators making use of their fast response times (tens of nanoseconds) and very high sensitivity [9, 10]. Cherenkov detectors are useful where there are sources of background ionising radiation as the threshold for light production is above typical Compton-electron energies produced by X-rays or synchrotron radiation. Such monitors, using the Cherenkov radiation produced in the glass of a photomultiplier tube (PMT), have been used on CEBAF at the Thomas Jefferson National Accelerator Facility [11]. Although relatively radiation tolerant, the darkening of the PMT glass due to ionising radiation has to be compensated by an increase of the PMT gain with time. The main advantage of PMT based detectors is their fast response time, enabling loss measurements on a bunch-by-bunch basis, which is of vital importance in (superconducting) Linacs with long bunch trains.

**Optical fibre detectors** Cherenkov light created in long optical fibres can be used to determine the longitudinal position of beam losses in a similar fashion to the long ionization chamber, with the fast response of the Cherenkov signal detected using photomultipliers at the end of the fibres [12]. Measurement of the light arrival time provides information on the position of the loss, while the integrated light amplitude is proportional to the size of the loss. A longitudinal

position resolution of 20 cm has been achieved in such systems. High purity quartz fibres (Suprasil ®) withstand radiation doses of up to 300 MGy and generate no scintillation photons, while scintillating fibres are  $\sim$ 1000 times more sensitive but not very radiation resistant [13].

## References

- [1] R.E. Shafer, Proc. 10th Beam Instrumentation Workshop (2002), p.44
- [2] W. Panofsky, SLAC Internal Report TN-63-57 (1963)
- [3] R.E. Shafer et al, Proc. 12th Int. Conf. on High Energy Accelerators FNAL (1983), p.609
- [4] D. Kramer et al, NSS '07. IEEE (2007) p.2327
- [5] E.B. Holzer, NSS '05. IEEE (2005) p.1052
- [6] S. Schlögl et al, Proc. 15th Int. Conf. on High Energy Accel, Hamburg (1992), p.254
- [7] V. Agoritas et al, CERN MPS/CO Note 71-51 (1971)
- [8] L. Fröhlich, Linac06, p.262
- [9] J.R. Parker et al, IEEE Trans. on Nucl. Science Vol. 18 No. 3 (1971), p.825
- [10] R.A. Lundy et al, IEEE Trans. on Nucl. Science Vol. 20 No. 3 (1973), p.596
- [11] J. Perry et al, PAC93, p.2184
- [12] F. Wulf, Proc. 9th European Workshop on Beam Diagnostics and Instrumentation for Particle Accelerators (2009), p.411
- [13] T. Kawakubo et al, EPAC'0, p.2652

#### 7.4.17 Schottky Spectra

##### 7.4.17.1 Transverse Schottky spectra and beam transfer functions

*O. Boine-Frankenheim, V. Kornilov,  
S. Paret*

Transverse Schottky signals and Beam Transfer Functions (BTFs) play an important role in the diagnostics of particle beams in storage rings and synchrotrons. The Schottky 'natural' noise spectrum, divided into separate bands, contains information on the incoherent and coherent frequencies in the beam [1]. If the beam is excited by an external signal the response gives the BTF which in addition provides the coherent stability threshold [2]. In many machines the Schottky spectrum as well as BTFs are routinely monitored to measure e.g. betatron tunes, tune spreads and stability properties of the beam.

The transverse Schottky signal of a beam is generated by the beam's dipole moment  $d(t) = \sum_{j=1}^N x_j(t) I_j(t)$  where  $x_j(t)$  is the offset and

$I_j(t)$  the current of the  $j^{th}$  particle at the position of a pick-up (PU) in the ring. The sum extends over all  $N$  particles in the detector. The Schottky noise power spectrum as a function of the frequency  $f$  is defined as  $S(f) = |d(f)|^2$  where  $d(f)$  is the Fourier transformed PU signal. If the beam is excited by an external amplitude spectrum  $G(f)$  the transverse BTF is defined as  $r(f) = d(f)/G(f)$ .

**Coasting beams** In coasting beams the transverse Schottky bands are centered around the betatron sidebands at the frequencies

$$f_n = Q_n f_0 \quad (1)$$

with  $Q_n = (n \pm Q_0)$ , the bare tune  $Q_0$ , the harmonic number  $n$  and the revolution frequency  $f_0$ . In the absence of intensity effects the shape of each band  $P_n^0(f)$  reflects the beam's momentum distribution. The width of a band

$$\delta f_n = \delta Q_n f_0, \quad (2)$$

is determined by the chromatic tune spread  $\delta Q_n = \xi_{\text{eff}} \delta_{\text{rms}}$  with the rms momentum spread  $\delta_{\text{rms}}$ , the effective chromaticity  $\xi_{\text{eff}} = |\pm\xi - \eta_0(n \pm Q)|$ , the chromaticity  $\xi$  and the slip factor  $\eta_0$ .

The BTF for low beam intensities is given through

$$r^0(z) = \int_{-\infty}^{\infty} \frac{P^0(z')}{z - z'} dz' \quad (3)$$

where  $z = (f/f_0 - Q_n)/\delta Q_n$  is the normalized frequency and  $P^0(z)$  is the low intensity Schottky spectrum. For a Gaussian distribution we obtain

$$r^0(z) = i \sqrt{\frac{\pi}{2}} w\left(z/\sqrt{2}\right) \quad (4)$$

where  $w$  is the complex error function. Other distributions are discussed in Ref.[3].

At high intensities the effect of impedances on the BTF can be described in terms of the real intensity parameters

$$\Delta U + i\Delta V = \quad (5)$$

$$\frac{r_p Z^2 N}{2\pi Z_0 A \gamma Q \delta Q_n} [\text{Im}(Z_{\perp}) + i\text{Re}(Z_{\perp})]$$

where  $Z_{\perp}$  is the transverse impedance at the frequency  $Q_n f_0$ ,  $Z_0$  is the vacuum impedance,  $r_p$  is the classical proton radius,  $A$  is the mass number and  $Z$  the charge state. The impact of direct space charge, inducing an incoherent tune shift  $Q_0 - \Delta Q_{sc}$ , is expressed by means of the space charge parameter

$$\Delta U_{sc} = \frac{\Delta Q_{sc}}{\delta Q_n}. \quad (6)$$

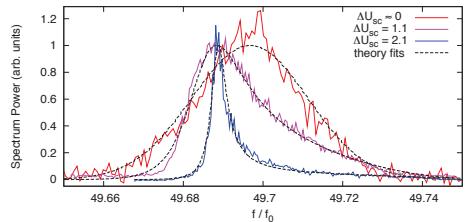


Figure 1: Lower Schottky bands at different beam intensities, measured in the heavy ion synchrotron SIS18 at GSI Darmstadt. The dashed lines are fits using the model Eq.(7) for the space charge effect. The legend gives the space charge parameters obtained from the fits.

The Schottky spectrum including coherent and incoherent intensity effects to first order is [4, 5]

$$P(z) = \quad (7)$$

$$\frac{P^0(z - \Delta U_{sc})}{|1 - (\Delta U + i\Delta V - \Delta U_{sc})r^0(z - \Delta U_{sc})|^2}$$

and the BTF

$$r(z) = \quad (8)$$

$$\frac{r^0(z - \Delta U_{sc})}{|1 - (\Delta U + i\Delta V - \Delta U_{sc})r^0(z - \Delta U_{sc})|}.$$

In Figs.1 and 2 Schottky side-bands and BTFs, respectively, measured for different intensities in a coasting ion beam are shown. The analytic formulas Eq.(7) and Eq.(8) are fitted to the measured results [5]. The space charge tune shift  $\Delta Q_{sc}$  can be obtained from the fit to the measured Schottky band or BTF. For a well known machine tune the analysis of the position and deformation of the signals allows the distinction of direct space charge and image current effects. A higher accuracy is achieved by supplying the momentum spread from an independent measurement. However, Schottky and BTF signals at high intensity are prone to perturbations and error sources.

For an analysis of the beam stability, a stability diagram can be produced from the BTF data [3]. The stability diagram is a plot of the imaginary part versus real part of the inverse BTF and allows the visualization of the stability limits due to Landau damping. A self-consistent analytical approach for transverse Schottky noise can be found in Ref.[6].

**Bunched beams** For a low intensity bunched beam the transverse Schottky band splits into equidistant synchrotron satellites

$$Q_{n,k} = (n \pm Q_0) + \Delta Q_k \quad (9)$$

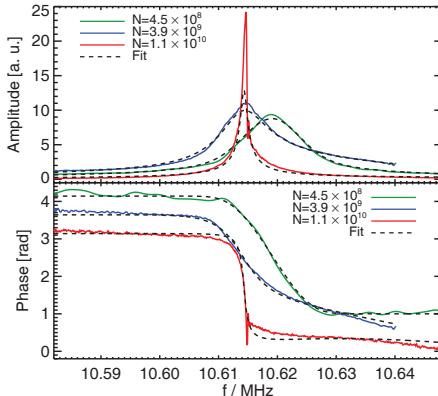


Figure 2: BTF signals measured with the same settings as the Schottky bands in Fig.1. The legend gives the actual number of ions in the ring.

where  $\Delta Q_k = \pm kQ_s$  and the synchrotron tune is  $Q_s$ . For a single particle performing betatron and synchrotron oscillations the relative amplitudes of the satellites are [1]

$$|d_k| \sim |J_k[((n \pm Q_0) - Q_\xi)\phi_m]| \quad (10)$$

where  $\phi_m$  is the longitudinal oscillation amplitude and  $Q_\xi = \xi/\eta_0$  is the chromatic tune.  $J_k$  are the Bessel functions of order  $k$ . For  $Q_n = Q_\xi$  only the central line ( $k = 0$ ) is important. The single particle spectrum is shown in Fig.3. In the absence of transverse nonlinear field components the width of each satellite with  $k \neq 0$  is determined by the synchrotron tune spread  $\delta Q_k \approx |k|Q_s\phi_m^2/16$ . For a given sideband the relative amplitude of the synchrotron satellites resembles the momentum distribution in the bunch. Therefore the envelope of the power spectrum duplicates the power spectrum  $S(f)$  for the equivalent coasting beam with the same rms momentum spread. (see Fig.4).

In the presence of incoherent space charge, represented by the tune shift  $\Delta Q_{sc}$ , or image currents effects, represented by a real coherent tune shift  $\Delta Q_{coh}$ , the shift of the Schottky satellites can be reproduced rather well by [7]

$$\Delta Q_k = -\frac{\Delta Q_{sc} + \Delta Q_{coh}}{2} \pm \sqrt{(\Delta Q_{sc} - \Delta Q_{coh})^2/4 + (kQ_s)^2} \quad (11)$$

which is the exact expression for the head-tail mode frequencies for an airbag bunch distribution in a barrier potential [9]. The above expression

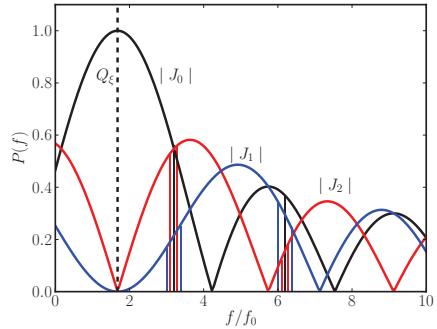


Figure 3: Single particle oscillation spectrum in a bunch. Shown are the satellites corresponding to two upper sidebands with  $n + Q_0$ .

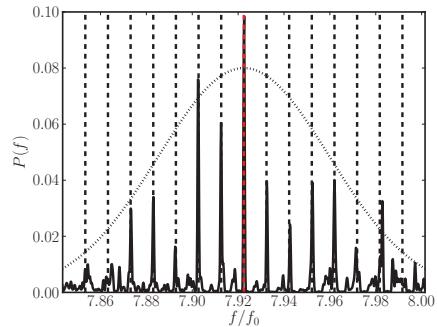


Figure 4: Synchrotron satellites in a Gaussian bunch obtained from a particle tracking code [7]. The dotted curve corresponds to the Schottky band for the corresponding coasting beam. Measured Schottky spectra from bunched beams are shown e.g. in [8].

can be used to estimate the space charge tune shift from the Schottky spectrum of the bunch.

## References

- [1] S. Chattopadhyay, Report 84-11, CERN, 1984
- [2] J. Borer, PAC79 p.3405
- [3] A. Chao, Physics of Collective Beam Instabilities in High Energy Accelerators, Wiley (1993)
- [4] S. Paret, O. Boine-Frankenheim, V. Kornilov, PRST AB 11 (2008) 7
- [5] S. Paret et al, PRST AB 13 (2010) 022802
- [6] D.V. Pestrikov, NIM A 578 (2007) 65
- [7] O. Boine-Frankenheim, V. Kornilov, PRST AB 12 (2009) 114201
- [8] T. Linnear and W. Scandale, PAC81 p.2147
- [9] M. Blaskiewicz, PRST AB 4 (1998) 044201

### 7.4.17.2 Longitudinal Schottky spectra and beam transfer function

*E. Shaposhnikova CERN*

The Schottky noise is a non-destructive diagnostic tool providing information about particle oscillation frequencies in the beam (see for example [1, 2, 3, 4]) The Schottky current  $\Delta I(t)$  is a sum of single particle signals  $i(t) = e\dot{\theta}\delta(\theta(t)-\theta_{PU}-2\pi k)$  generated each  $k$ -th passage of the pick-up (PU) at azimuthal position  $\theta_{PU}$ . Assuming  $\Delta I(t)$  describes a stationary stochastic process, the spectral density of the Schottky signal is

$$P(\omega) = \int_{-\infty}^{\infty} \langle \Delta I(t) \Delta I^*(t - \tau) \rangle e^{i\omega\tau} d\tau. \quad (12)$$

**Unbunched beam** For the  $n$ -th particle with revolution frequency  $\omega_{0n}$  and phase  $\omega_{0nt} + \theta_n$

$$i_n(t) = \frac{e}{2\pi} \sum_{k=-\infty}^{\infty} \omega_{0n} e^{-ik(\omega_{0nt} + \theta_n)}, \quad (13)$$

where  $\theta_n$  is the initial ( $t=0$ ) phase relative to the PU. After summing the contribution from all  $N$  particles, averaging over random  $\theta_n$  and subtracting the DC ( $k=0$ ) beam current

$$P(\omega) = \frac{e^2}{2\pi} \sum_{n=1}^N \sum_{\substack{k=-\infty \\ k \neq 0}}^{\infty} \omega_{0n}^2 \delta(\omega - k\omega_{0n}). \quad (14)$$

Using the distribution function in revolution frequency  $F(\omega_0) = dN/d\omega_0$ , normalised to 1 and with a spread  $\delta\omega_0$  much smaller than the average value  $\bar{\omega}_0$ , the spectral density of the Schottky signal is

$$P(\omega) = \frac{e^2 N \bar{\omega}_0^2}{2\pi} \sum_{\substack{k=-\infty \\ k \neq 0}}^{\infty} \frac{1}{|k||\varepsilon_k(\omega)|^2} F\left(\frac{\omega}{k}\right), \quad (15)$$

where  $\varepsilon_k(\omega) = 1$  for low intensity and  $\varepsilon_k(\omega) = 0$  describes an excitation of coherent modes due to impedance, see next section (BTF). This signal has lines at all revolution harmonics  $k$  with increasing width  $k\delta\omega_0$  and power density decreasing as  $1/k$ . For non-overlapping bands each multipole represents a particle distribution in revolution frequencies and therefore in momentum. For large space charge below transition (or inductive impedance  $Z(\omega)$  above) the amplitude of the Schottky band has two clear peaks separated by a distance  $\propto \sqrt{N|Z(\omega)|}$  and its variation with intensity can be used for impedance estimation.

**Bunched beam** For a bunched beam  $\theta(t) = \omega_0 t + \phi(J, \psi(t))/h$ , where  $\phi$  is the RF phase,  $J$  and  $\psi(t) = \omega_{st} t + \psi_{0n}$  are action and angle variables. The spectral density is

$$P(\omega) = \frac{e^2 N \omega_0^2}{2\pi} \times \sum_{k=-\infty}^{\infty} \sum_{\substack{m=-\infty \\ m \neq 0}}^{\infty} \frac{1}{|m|} F\left(\frac{\omega - k\omega_0}{m}\right) |I_{mk}(J)|^2, \quad (16)$$

where  $F(\omega_s) = dN/d\omega_s$  is a distribution function in synchrotron frequency  $\omega_s$  and function

$$I_{mk}(J) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i\phi(J, \psi)/h - im\psi} d\psi \quad (17)$$

is taken at the value of  $J$  corresponding to the synchrotron frequency  $\omega_{s,mk} = (\omega - k\omega_0)/m$ . The expression (16) is valid for an arbitrary RF system. For short bunches in a single RF system without acceleration  $I_{mk} \simeq i^m J_m(k\phi_a/h)$ , where  $J_m(x)$  is an ordinary Bessel function of integer order  $m$  and  $\phi_a$  is an RF phase amplitude corresponding to  $\omega_{s,mk}$ . Each  $k$ -th revolution band is split into an infinite number of synchrotron satellites peaked at  $k\omega_0 \pm m\omega_{s0}$ , where  $\omega_{s0} = 2\pi/T_{s0}$  is a linear synchrotron frequency. For large  $k$  the revolution band has a width defined by  $k\delta\omega_0$  and describes the particle distribution in revolution frequencies and therefore momentum. The current in the central line  $k\omega_0$  is proportional to  $N$  and in satellites to  $\sqrt{N}$  with increasing width  $m\delta\omega_s$ . The shape of each synchrotron sideband is determined by the particle distribution in synchrotron frequency (with width  $\delta\omega_s$ ), but is significantly modified by the formfactor  $I_{mk}^2$ , Figs.5-6. The incoherent synchrotron frequency shift measured from the distance between two sidebands (at large  $m$ ) as a function of intensity can be used to estimate the longitudinal effective reactive impedance of the ring, see example in [6].

To increase sensitivity and minimise coherent signals measurements are usually done with narrow-band (resonant) PU at frequencies much higher than the RF frequency  $\omega_{rf} = \hbar\omega_0$ , but where Schottky bands still don't overlap. There is no fundamental difficulty in providing a pickup with sufficient sensitivity to observe Schottky signals. The choice of low-noise and high dynamic range front-end amplifier, the signal processing method and spectrum analyser depend on signal characteristics. A realistically attainable noise level for the complete electronics chain is 5-10

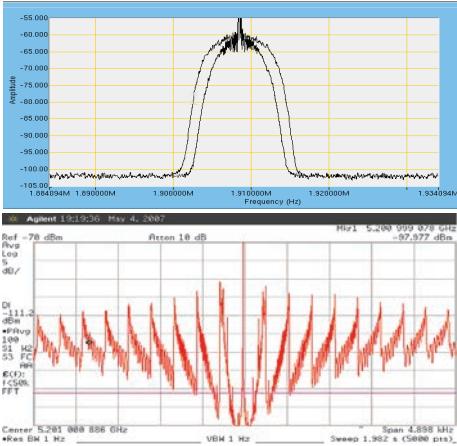


Figure 5: Longitudinal Schottky spectrum of a bunched beam at 5.2 GHz measured before and after momentum cooling in RHIC (top, 50 kHz span). Bottom: wide span of top signal after cooling. All in logarithmic scale (dB),  $\omega_0/(2\pi) = 78$  kHz,  $\omega_{s0}/(2\pi) = 270$  Hz[5].

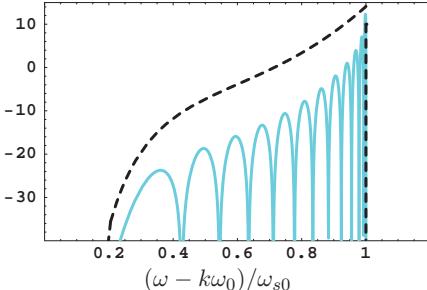


Figure 6: Dipole ( $m=1$ ) sideband (a.u.) in the longitudinal bunched beam Schottky spectrum at  $k = 12h$  (the LHC 4.8 GHz system) for infinite acquisition time  $T_a$ , and corresponding distribution function  $2\omega_{s0}F(\omega - k\omega_0)$  (dashed line) for a Gaussian bunch with  $\sigma_\phi = \pi/4$ . All in logarithmic scale (dB).

dB. Care must be taken with processing technique (filtering, down-mixing, ...) to obtain sufficient dynamic range in the presence of the strong coherent signals. For weak single bunches higher signal to noise ratio can be achieved by averaging and gating after high-gain pre-amplification. An instantaneous dynamic range of 100 dB has been obtained in LHC [7].

**PD Schottky spectrum** The peak detected (PD) spectrum is a special case of the bunched beam

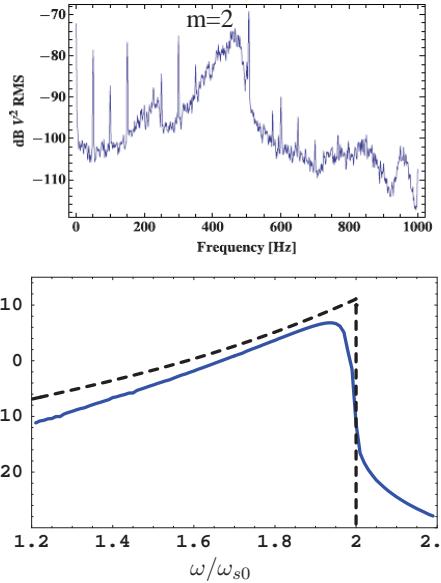


Figure 7: Longitudinal PD Schottky spectrum of bunched beam. Top: spectrum measured in the SPS for low intensity bunch with  $\sigma_\phi = \pi/4$ ,  $\omega_{s0}/(2\pi) = 240$  Hz. Bottom: quadrupole band  $P_U/P_0$  for  $\sigma_\phi = \pi/4$  and  $\Phi = \pi/8$ , measurement time  $T_a = 320 T_{s0}$ ;  $\omega_{s0}F(\omega/2)$  as dashed line. All in logarithmic scale (dB).

longitudinal Schottky diagnostics developed and used at CERN (SPS, LHC) since the late eighties. The PD signal measured each revolution turn at  $t = t_k$

$$U_k = \Delta U_k + U_{k-1} e^{-\delta} \quad (18)$$

is proportional to the average bunch peak amplitude

$$\Delta U_k = \frac{B}{2\Phi} \int_{-\Phi}^{\Phi} I(t_k, \phi) e^{\alpha\phi} d\phi \quad (19)$$

with parameters  $B$ ,  $\Phi$ ,  $\delta$  and  $\alpha$  defined by the experimental set-up (wide-band PU, fast diode and capacitor) [8]. The spectral density of the quadrupole line of the PD signal

$$P_U(\omega) = \frac{P_0}{\omega_{s0}^2} \times \sum_{m=1}^{\infty} \int \Omega^2 F(\Omega) |A_m(\Omega)|^2 |S(\omega - m\Omega)|^2 d\Omega, \quad (20)$$

where  $P_0 = e^2 N f_{s0} B^2$ , well represents the particle distribution in synchrotron frequency  $F(\Omega)$ , Fig.7, affected largely only by the formfactor

$|A_m|^2$ . For short bunches in a single RF system  $A_2 = \frac{1}{3}(\frac{\phi_a}{\Phi})$  for  $\phi_a \leq \Phi$  and  $A_2 = 1 - \frac{2}{3}(\frac{\Phi}{\phi_a})^2$  for  $\phi_a \geq \Phi$ . All odd multipoles are suppressed in amplitude with  $A_1/A_2 < 10^{-1}$  and high even multipoles are very distorted. The shape of the quadrupole line can be used to see details of the distribution including holes due to depletion by noise excitation as well as different coherent lines. The fact that measurements are done during an acquisition time  $T_a$  at some sampling rate  $t_s$  is described in (20) by the function

$$\begin{aligned} |S(\omega - m\Omega)|^2 = \\ \frac{2t_s^2}{T_a} \frac{\sin^2[(\omega - m\Omega)T_a/2]}{\sin^2[(\omega - m\Omega)t_s/2]}. \end{aligned} \quad (21)$$

This function, replacing the  $\delta$ -function in frequency used to obtain (15-16), broadens the measured spectrum lines and is the same for traditional Schottky measurements.

**Longitudinal beam transfer functions** The beam transfer function (BTF) relates the frequency domain response of the beam to a known excitation (e.g. white noise or swept signal in the band of interest). (see for example [9, 10, 11]) Therefore, compared to Schottky measurements, one more element, a longitudinal kicker, is required. The output is a noise signal and similar processing techniques to Schottky can be applied. For bunched beam these measurements can also be done at frequencies close to  $\omega_{rf}$  using RF phase or amplitude modulation at synchrotron frequency multiples, but the excitation amplitude should be small enough to minimise the effect on particle motion (e.g. creation of islands). BTFs give information about particle distribution and are widely used to characterize beam control systems and for beam stability analysis.

For low intensity unbunched beams BTFs,  $B_k^0(\omega)$ , are described in [12]. Their measurement gives the most direct information about the stability diagram (inverse BTF). For high intensities, the azimuthal harmonic of the beam current perturbation is connected to voltage perturbation  $\Delta V_k(\omega)$  by

$$\Delta I_k(\omega) = B_k^0(\omega)/\varepsilon_k(\omega) \Delta V_k(\omega), \quad (22)$$

where  $\varepsilon_k(\omega) \simeq 1 + iB_k^0(\omega)Z(\omega)/(2\pi k)$  and BTF measurements at different intensities can give a value of beam coupling impedance  $Z(\omega)$ . For large space charge impedance below transition (or inductive above) the amplitude of the BTF has two peaks similar to the Schottky signal.

Unlike the unbunched beam case, bunch excitation at  $k\omega_0 + \Omega$  results in a response at different frequencies  $p\omega_0 + \Omega$

$$\Delta I_p(\Omega) = \sum_{k=-\infty, k \neq 0}^{\infty} G_{pk}(\Omega) \Delta V_k(\Omega), \quad (23)$$

described by beam transfer matrices

$$G_{pk}(\Omega) = -\frac{2iI_0h}{V \cos \phi_s} \sum_{m=1}^{\infty} \frac{m^2}{k} g_m^{pk}(\Omega), \quad (24)$$

where  $I_0$  is the average beam current,  $\phi_s$  is synchronous phase and  $V$  is the RF voltage amplitude. For  $M$  identical and equidistant bunches  $G_{pk}(\Omega) \neq 0$  only if  $p = k + lM$ ,  $l = 0, 1, \dots$ . The matrix element  $g_m^{pk}(\Omega)$  is proportional to the dispersion integral

$$\begin{aligned} g_m^{pk}(\Omega) = \\ \int_0^\infty \frac{dF(J)}{dJ} \frac{I_{mp}(J)I_{mk}^*(J)\omega_s(J)dJ}{(\Omega - i\sigma)^2 - m^2\omega_s^2(J)}, \end{aligned} \quad (25)$$

where the integration contour is chosen to satisfy the initial conditions,  $F(J) = dN/dJ$  is the particle distribution function and  $I_{mk}(J)$  is defined by (17). For dependence  $\omega_s(J)$  inside the bunch such that  $\omega'_s(J_0) = 0$  only if  $dF/dJ = 0$ ,

$$\begin{aligned} g_m^{pk}(\Omega) = \\ \mathcal{P} \int_0^\infty \frac{dF}{dJ} \frac{I_{mp}(J)I_{mk}^*(J)\omega_s(J)dJ}{\Omega^2 - m^2\omega_s^2(J)} \\ + i \frac{\pi}{2m^2} \frac{dF}{dJ} \Big|_{J=J_0} \frac{I_{mp}(J_0)I_{mk}^*(J_0)}{|\omega'_s(J_0)|}, \end{aligned} \quad (26)$$

where  $\mathcal{P}$  is the principal value of the integral and  $J_0$  satisfies the resonant condition  $\Omega = m\omega_s(J_0)$ .

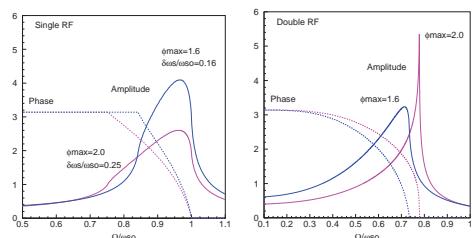


Figure 8: Amplitude (a.u.) and phase (rad) of the longitudinal BTF element  $g_1^{11}(\omega)$  for small (blue) and large (red) emittance bunch in a single (left) and in a double RF system (right) with ratio two of harmonics and voltages, anti-phase,  $\omega_s(0) = 0$ . Binomial ( $\mu = 2$ ) distribution function in particle energy.

The case where  $\omega'_s(J) = 0$  for some  $J \neq 0$  inside the bunch exists in a double RF system and is also possible in a single RF system for certain particle distributions due to potential well distortion. Landau damping in this region is lost, this appears as a spike in the BTF (Fig.8, right) and can perturb the operation of the beam control system. Change in the BTF phase slope indicates minimum and maximum synchrotron frequency in the bunch and can be used for measurements of the incoherent frequency spread as well as its shift with intensity and therefore effective inductive impedance. The shape of BTF depends on particle distribution. For a symmetric potential well (no acceleration and energy loss) only even terms are present in the beam response for amplitude modulation and only odd terms for phase modulation.

## References

- [1] J. Borer et al, PAC'74, CERN/ISR-DI/RF/74-23
- [2] S. Chattopadhyay, Report 84-11, CERN, 1984
- [3] D. Boussard, CAS, 1985, CERN-87-03, p.416
- [4] S. Berg, CERN SL-97-49, 7.12.1999
- [5] J. M. Brennan et al, EPAC02, p.308
- [6] M. Blaskiewicz et al, EPAC'02, p.1488
- [7] F. Caspers et al, PAC07, p.4174
- [8] E. Shaposhnikova, T. Bohl, T. Linnecar, Proceed. HB'2010
- [9] J. Borer et al, IEEE Trans. Nucl. Sci., NS-26, 1979, p.3405
- [10] A. Hofmann, B. Zotter, IEEE Trans. Nucl. Sci., NS-24, 1977, p.1487
- [11] E. Shaposhnikova, CERN SL/94-19 (RF), 1994
- [12] A. Chao, B. Zotter in Handbook of Accelerator Physics and Engineering, World Scientific (2006) 3rd print, p.139

## 7.5 IMPEDANCE DETERMINATION

### 7.5.1 Bench Measurements

*F. Caspers, CERN*

The standard way to do bench measurements of the beam coupling impedance [1] on particle accelerator components is the coaxial wire method [2, 3, 4, 5]. The technique was proposed in [2]. It assumes that a bunch of a highly relativistic beam has a very similar EM field distribution (TEM field) to a short pulse on a coaxial line. For a discussion of measurement validity, see [6].

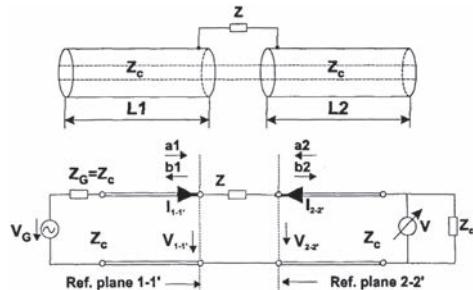


Figure 1:  $S$ -parameters, lumped impedance.

Consider a single lumped impedance  $Z$  in some coaxial line with a characteristic impedance  $Z_C$ . The common technique to measure such an impedance by means of modern instrumentation in the rf and microwave range is the use of a VNA (vector network-analyser) [7, 8, 9]. As a primary result we obtain  $S$ -parameters ( $S$  = scattering) which describe the complex ratio of two waves. These two waves are denoted as "a" when incident to the DUT (= device under test) and "b" when travelling away from the DUT (Fig.1 [9]). A wave "b" in general is the superposition of a reflected part from and a transmitted part through the DUT. A generator with some voltage  $V_G$  and a source impedance  $Z_G = Z_C$  will launch into a properly terminated line with characteristic impedance  $Z_C$  a forward travelling wave with half the amplitude of the generator voltage  $V_G$ . In terms of voltages, "a" can be defined then as  $a = V_G/(2\sqrt{Z_C})$ , thus the voltage related only to the forward travelling wave on the line, normalized to the square root of the line impedance. The (average) power transported by that wave is  $P = aa^*$  if  $V_G$  stands for the rms and  $P = 0.5 aa^*$  if  $V_G$  is the peak value. The voltage (e.g.  $V_{1-1'}$ ) and current in a certain reference plane (1-1'), c.f. Fig.1, results from the superposition of the two waves a and b. The forward transmission coefficient (from port 1 to port 2) of the DUT is denoted here as  $S_{21}$  (not  $S_{12}$  due to the way the  $S$ -matrix is defined [7]) and the input reflection coefficient seen at port 1 is  $S_{11}$  [7, 8]. We assume that our set-up is calibrated i.e. that for  $Z = 0$  we obtain  $S_{21} = b_2/a_1 = 1$ . The exact relation between the  $S$ -parameter (obtained from the VNA) and the unknown (complex) impedance  $Z$  is given as [4, 9]

$$Z = 2Z_C(1 - S_{21})/S_{21} \quad (1)$$

This is the same formula which can be used via a software button on certain modern instruments (VNAs) in order to directly display  $Z$  and not  $S_{21}$  [7]. In principle,  $Z$  may also be deduced from a measured  $S_{11}$  (reflection measurement), but for several good reasons such as difficulties in defining a reference plane [4] and mutual interaction of reflected waves from several discontinuities, this technique is not very suitable for the present problem and should be avoided. The concept shown so far is only valid for a single, lumped impedance which is geometrically smaller than the shortest wavelength under consideration and whose length is less than the diameter of the coaxial line (beam-pipe). In particular any error between electrical and mechanical length (electrical delay correction) is attributed to the impedance and may lead to unphysical results such as a negative real part [10].

When doing practical bench measurements, we normally compare a reference measurement with a DUT result. In this case we replace  $S_{21}$  by the normalized value

$$S_{21,N} \equiv \frac{S_{21,\text{DUT}}}{S_{21,\text{REF}}} \implies Z = 2Z_C \frac{(1 - S_{21,N})}{S_{21,N}} \quad (2)$$

Eq.(2) is still exact in the single lumped impedance frame and should not be confused with a similar looking Eq.(3) given as an approximation (for  $Z \ll Z_C$ ) in a different notation [4],

$$Z = 2Z_C (S_{21,\text{REF}} - S_{21,\text{DUT}}) / S_{21,\text{REF}} \quad (3)$$

Early work [3, 5] employed excitation of the REF and DUT in parallel through a splitter, the outputs being passively subtracted, and Eq.(3) being applied to obtain  $Z$ . With a pulse generator and sampling scope as detector, good loss factor measurements were made. However, many error sources inherent in that method have been eliminated through sequential measurements of REF and DUT using a modern VNA. The latter is now to be recommended.

**Loss factor** (see Sec.3.2.6) The loss factor also requires a (time domain trace) reference data set and a DUT data set. Those data may be obtained from frequency domain data via inverse FFT (Fast Fourier Transform) or a similar algorithm (e.g. CHIRP-Z transform) [7]. Note that “low pass” type data acquisition in the frequency domain is mandatory (i.e. all discrete frequencies are equidistantly spaced with respect to 0 or DC and not just with any arbitrary frequency band

(“band-pass mode”). Time domain data acquisition with a sampling scope may be employed but it is usually less precise than the synthetic pulse method now available [9, 10]. A notation similar to [3] has been kept here [Eq.(4)], in order to show the convertibility from frequency to time domain data (and vice versa) which is always possible for linear, time invariant networks. Eq.(4) is expressed in terms of  $S_{21}(t)$  parameters via IFFT from  $S_{21}(\omega)$ . This is equivalent to earlier work [3] using sampling scopes for time domain measurements in terms of current  $I(t)$ , which corresponds now to  $S_{21}(t)$ ,

$$k(\sigma) = 2Z_{C,\text{REF}} \times \frac{\int S_{21,\text{REF}}(t)[S_{21,\text{REF}}(t) - S_{21,\text{DUT}}(t)]dt}{\int S_{21,\text{REF}}^2(t)dt} \quad (4)$$

**Measurement accuracy** The measurement accuracy which can typically be obtained using equipment like sampling scopes and VNAs is limited by the connector reproducibility which amounts  $\pm 0.05$  dB for good devices. The absolute attenuation reading of a modern VNA has the same error bar, in the range 0-20 dB for frequencies up to 20 GHz. Timing stability after instrument warm-up is about 1ps over 30 minutes (usually not specified this way). Proper calibration procedures [7, 8] permit nearly perfect elimination of generator and load mismatch, including connecting cables. These numbers indicate an absolute reading accuracy for  $Z$  of about 1% when the value of the DUT is in the order of  $Z_C$ . For smaller values of  $Z$  the limitation given by the 0.05 dB mentioned above, coincides with the VNA trace stability (vertical drift). This trace stability may be considerably worse for sampling scopes. Usually one is not looking at values  $Z \gg Z_C$  as this is beyond the limitation of the wire method itself. However in such a case the instrument (VNA) has still an absolute precision within  $\pm 5\%$ .

**Transverse impedance** The situation is more involved as two components of the transverse impedance in the linear regime have to be determined to correctly describe the transverse beam dynamics: (1) the dipolar (or driving) impedance, which is related to the displacement of the leading particle, and (2) the quadrupolar (or detuning) impedance, which is related to displacement of the trailing particle [18, 19]. Note that in the case of an ultrarelativistic beam and symmetric structures, the quadrupolar term disappears: this

complexity appears only when the structure is asymmetric or when the particles are not ultrarelativistic [20] (this latter case will not be discussed below).

**Dipolar transverse impedance** The standard technique is to stretch two parallel wires across the DUT (“two wire method”). For asymmetric structures, the measurement must be done both in the horizontal and in the vertical plane. These two wires form a “shielded pair” transmission line and this structure may carry odd (+−) and even (++) modes. For the transverse measurement we are only interested in the odd mode excitation and the common mode should be suppressed as well as possible. For this purpose one should use a very broadband coaxial 180° hybrid (e.g. from 2 MHz to beyond 2 GHz) having a suppression (= isolation) of the undesired mode by >30 dB over the full frequency range. Having such a hybrid at either end of the DUT permits common mode suppression by >60 dB. The lateral distance  $\Delta$  of the two wires should not exceed the radius of the DUT or beam pipe (for the reference measurement). Usually  $\Delta$  is rather small (~10% of the radius of the beampipe) in order to stay within a regime of weak perturbation of the fringe field of the two-wire line. This parallel wire transmission line has for each mode (common or even mode and odd mode) a well defined characteristic impedance  $Z_{C,ODD}$  and  $Z_{C,EVEN}$  which can be determined from its geometry. In the case of two conductors which are reasonably close to each other, the characteristic impedance may be found using the analysis of a two-wire line in free space, since there is very little influence of the shielding beampipe. The odd mode characteristic impedance is often selected to be 100 Ω in order to be matched (50 Ω + 50 Ω) to the output of the 180° hybrid (Fig.2 [9, 10]). Applying Eq.(1) and simply replacing  $Z_C$  in Eq.(1) by  $Z_{C,ODD}$  returns

from the measured  $S$ -parameters a value for  $Z$ . This  $Z$  then yields the transverse impedance [10]

$$Z_{TR} = \frac{cZ}{\omega\Delta^2} \quad (5)$$

As a particular method for transverse dipolar impedance the image plane approach [9] must be noted as well. Finally, for the evaluation of low frequency transverse impedances where the two-wire method lacks sensitivity, the “probe coil technique” is more suitable [21].

**Quadrupolar transverse impedance** To assess the quadrupolar impedance, another method using a single wire displaced transversally, known as the “displaced single wire method”, is needed [22]. With the two wires (or single wire on image plane), only the dipolar impedance could be evaluated. With a displaced single wire, the “generalized impedance”, defined as the sum (or difference depending on the plane, see formula (2) in [5][B]) of the dipolar and quadrupolar impedance, is obtained. Combining the two wire and displaced single wire techniques, the dipolar and quadrupolar contributions can be disentangled. Note that the quadrupolar impedance, which is null in symmetric structures, can be smaller than, equal to or larger than the dipolar impedance. An example of the latter case is given for the case of kickers[23, 24, 25].

**Distributed impedances** If the physical length of the DUT is larger than the beam pipe diameter, the impedance is not “lumped” but is “distributed”. Loop coupler type kickers are examples of distributed impedances. Application of the lumped impedance analysis could result in an unphysical, negative real part of the impedance. Consequently it is recommended that one apply [11] the “log” formula Eqs.(8),(9). Note that measurements on high voltage (HV) kickers are, in particular, delicate since the length and termination of the coaxial cables connecting the individual modules to the HV pulse generator play an important role, as well as the presence of the kicker tanks which may form a kind of cavity [11]. The correct translation of the change in  $S$ -parameter from the wire measurement into meaningful beam coupling impedance, may require sophisticated modeling [11] of the DUT, and also shows the limits of the wire method which is basically a perturbation approach.

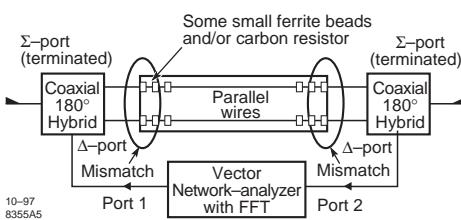


Figure 2:  $Z_{TR}$  impedance measurement.

Consider a coaxial transmission line (length  $l$ ) with the complex propagation constant  $\gamma$  determined by  $L'$ ,  $C'$ ,  $R'$  and  $G' = 0$  (no dielectric losses for the DUT). We get for a complex wall impedance per unit length  $\zeta$  and  $i = -j$  to be consistent with data from VNAs and Refs.[10, 11],

$$Z_{C,DUT} = Z_{C,REF} \sqrt{1 + i \frac{\zeta}{k_0 Z_{C,REF}}} \quad (6)$$

Here  $Z_0 = Z_{C,REF}$  equals  $\sqrt{L'/C'}$ ,  $k_0 = k_{REF} = \omega \sqrt{L'C'}$  and

$$k_{DUT} = k_0 \sqrt{1 + i \frac{\zeta}{k_0 Z_0}} \quad (7)$$

For  $|\zeta| \ll k_0 Z_{C,REF}$ , one obtains

$$S_{21,DUT}/S_{21,REF} = \exp[-i(k_{DUT} - k_0)l] \quad (8)$$

Using approximations quoted in [10], we find

$$Z = \zeta l \approx -2Z_0 \ln \left( \frac{S_{21,DUT}}{S_{21,REF}} \right) \quad (9)$$

The solution of Eq.(9) is not unambiguous and thus only the principal value of the complex logarithm should be taken.

Note that there may be 3 independent reasons to obtain a negative real part for the impedance when doing bench measurements: (i) Incorrect elimination of taper (end-piece) effects and/or mismatches of the generator or receiver ports (load); (ii) Incorrect electrical delay in the frame of the single lumped element model; (iii) Limit of validity of the single lumped element model (to be replaced by distributed impedance model or multiple lumped element model).

**General remarks** Note that the presence of the wire always lowers the  $Q$  of high  $Q$  DUTs like a cavity and usually detunes the resonance frequency (upwards) by a few % to  $>30\%$  (in some cases). The wire permits exchange of EM energy between discontinuities for frequencies below wave-guide cut-off which would not be the case without the presence of the wire. Certain (but not all) cavity-like objects may be measured in terms of loss factor or  $R/Q$  using the wire method. The  $R/Q$  is practically independent of  $Q$ , if without changing the shape of the cavity only the wall current losses are modified. For short gaps (gap length  $\ll$  beam-pipe diameter)

and a thin wire in a single cell cavity, a change in  $Q$  due to the wire does not (to first order) change the  $R/Q$  value. But there are cases known, where wire based  $R/Q$  measurements on cavities returned results that were off by  $>50\%$  with respect to the correct value confirmed by bead-pull or numerical simulation (see Sec.6.16). Thus wire based longitudinal impedance measurements on cavity like objects are possible, but the results should be handled cautiously.

An increase in diameter of the beam pipe (step out) represents to the beam, for lower frequencies, an inductive impedance. However for the wire measurement this situation would show up as a parallel capacitance (parallel between inner and outer conductor). Thus in terms of correct modeling this parallel  $C$  has to be converted into a shunt  $L$  for the  $S_{21}$  measurement and the model to be chosen is a series  $L$  [4, 9]. For the beam there is clearly a difference between a “step in” and a “step out”. The step out has always a positive loss factor (or positive real part) i.e. the beam loses energy on such a discontinuity, but the step in may have a negative loss factor (or negative real part) and the beam is gaining energy. This does not show up in the wire method since by definition  $S_{21}$  is equal to  $S_{12}$ , but when measuring a cascade of a step in and step out the sum of the loss factors of both discontinuities is given reasonably correctly. Note that the sum of the loss factors of a shallow taper in and taper out are close to zero, but both tapers together may form a cavity like object. This means in practice, that a loss-factor measurements which implies time-domain gating, is not sensitive to the exchange of energy between those two discontinuities, but in the frequency domain it may appear as a cavity (coupling between wave-guide cut-off, due to the wire). A common problem is the matching of the wire. As already mentioned, impedance tapers are not the ultimate solution (strong reflections at low frequencies) and radial tapers (with constant impedance) may require (unless  $50\Omega$  all the way) some broadband (resistive) matching device at both ends of the beam pipe. For transverse measurements, long impedance tapers are practically impossible. In this case one may simply consider using a flat end plate with some resistive matching network. Note that for resistive matching networks a good choice are tiny carbon resistors ( $0.05\text{-}0.1\text{ W}$ ) which can be used up to several GHz, rather than precision metal film resistors which often have a large inductance.

**Suggestions about the use of the measurement instrument** The most powerful tool to carry out bench type impedance measurements is the VNA with time domain option. An array of complex data points (usually  $S_{21}$  in equidistant frequency steps) taken in the frequency domain including a “frequency domain” cable and connector calibration [7, 8] is converted via a Fourier transform algorithm (nowadays often Chirp-Z type [7]) into equivalent time-domain data. For meaningful loss-factor evaluation only the so called “low pass” mode [7] of the instrument must be applied and not band-pass mode. The latter may be applied however for time-domain filtering (gating) of impedance data. There is a certain risk of under-sampling in the frequency domain in particular since certain VNAs start at 50 MHz and go beyond 20 GHz. In the low-pass mode the measurement frequencies would be e.g. 50, 100, 150 MHz ... which clearly has the risk of missing resonances and to run into aliasing problems (synthetic pulse distance 20 ns). This difficulty is sometimes referred to as the “strong reverberation problem” and means, in plain words, that all ringing of the DUT in the time-domain should be down to almost zero before the next (periodic) pulse arrives. Note that also the side-lobes of the synthetic pulse should be down to -60 dB before the next one appears (choice of proper weighting function) [7, 8]. The advantage of frequency domain measurement (as compared to sampling scope) is the better signal to noise ratio and the possibility to easily apply certain calibration techniques [8]. With a time and frequency domain trace available, both loss factor and impedance can be determined using a single instrument and set-up. The often recommended impedance tapers [3, 4, 5] are in reality not broadband transformers but rather behave like bad connectors with a lot of reflections at low frequencies. Their influence on the final result can be removed by time-domain gating and/or calibration and de-embedding procedures (TRL,TSD) [7, 8].

### Special methods

**Slow beam simulator** For test bench simulations of slow beams a quasi-static approach has been developed [12], that permits dialing virtually any  $v/c$  value (even  $>1$ ) as well as independently any  $E/H$  ratio. It essentially applies the image plane approach but instead of a strip-line, a row of closely spaced small probes and loops is provided. Each of those probes and loops is fed with an rf

signal from some broadband (few MHz to some GHz) power divider, which arrives at the probe or loop via a thin coaxial cable of well defined delay with respect to its neighbor probe or loop. The probes serve as  $E$ -field antennas and the loops as  $H$ -field antennas, but they are not used to radiate EM energy (near field use only).

#### Measurements beyond wave-guide cut-off

Already in the early days of loss factor measurements, the spectrum of the pulse had significant components beyond wave-guide cut-off [3, 5]. The hope was, that HOMs would not become visible and perturb the first arriving pulse. This approach can be justified to some extent using smooth tapers but in the frequency domain, HOMs clearly show up as notches in the transmission. Those notches can be reduced [13] by inserting microwave absorbing material (foam) near the tapers or end pieces of the test set-up. One can even go a step further and omit the wire completely, playing on the fact that certain TM modes have a similar surface current distribution as the TEM field [13]. The technique to operate without any wire limits of course the frequency range but permits wave-guide mode, pulse-reflectometry [14] to identify obstacles and band-limited impedance measurement in transmission.

#### Measurements of very small impedances

Very small impedances have been measured by Walling et al in a high precision “standard set-up” as well as converting the DUT into a TEM resonator [15]. A reliable technique is to apply an inner conductor, which is capacitively (weakly) coupled and operates as a transmission resonator. The weak capacitive coupling and measurement of the  $Q$  values and resonance frequencies in transmission, permits rather straightforward evaluation of the real and imaginary part of the beam-pipe impedance. The change in resonance frequency of this TEM line resonator (as compared to a reference measurement) indicates the imaginary part of the wall impedance (using the identical inner conductor), and a change in  $Q$  stands for the real part. The advantage of capacitive coupling at either end is that there will be no doubts on contact reproducibility. For measurements of loss changes when lowering the temperature of the DUT towards the cryo region (LHC) a technique using two parallel inner conductors (as their conductivity changes as well) has been developed [16]. Those two inner

conductors are resonating in odd and even mode thus permitting the independent evaluation of their own conductivity (as a function of temperature) and the conductivity of the DUT (= LHC liner section). Also in the context of liner measurements (SSC and LHC) the triaxial impedance measurement technique [17] has been developed to evaluate the impact of the (beam related) transmitted signal via the liner perforations.

## References

- [1] V.G. Vaccaro, CERN-ISR-RF 66-35 (1966)
- [2] A. Faltens et al, Proc. 8th Int. Conf. High Energy Acc. (1971), p.338
- [3] M. Sands, J. Rees, SLAC report PEP-95 (1974)
- [4] H. Hahn, F. Pedersen, BNL 50870, Particle Accelerators and High Voltage Machines -TID-4500 (1978)
- [5] M.G. Billing, J.L. Kirchgessner, R.M. Sundelin, PAC 79, p.3583
- [6] H. Hahn, PRST AB Vol. 3, 122001 (2000)
- [7] R.A. Witte, Spectrum and Network Measurements, PTR Prentice-Hall (1993)
- [8] G.H. Bryant, IEE Electrical Measurement Series 5, Peter Peregrinus Ltd (1993)
- [9] F. Caspers, Lecture Notes in Physics No 400, Springer (1992) p 80
- [10] L.S. Walling et al, NIM A281 (1989) p.433
- [11] H. Hahn, M. Morvillo, A. Ratti, BNL report AD/RHIC/RD-95 (1995); AD/RHIC/RD-105; AD/RHIC/RD-112
- [12] F. Caspers, D. Möhl, A. Schwinn, EPAC 90, p. 1037; CERN-PS 90-35
- [13] G. Lambertson et al, LBL 28190; EPAC90, p.1049
- [14] J. Jacob et al, EPAC92; CERN/PS 92-30
- [15] F. Caspers , T. Scholz, CERN-PS 95-09 (AR); PA Vol.50 No.123
- [16] F. Caspers, M. Morvillo, F. Ruggiero, PAC 97; CERN LHC report 115
- [17] F. Caspers, E. Jensen, F. Ruggiero, EPAC92; CERN/PS 92-24 (RF-AR)
- [18] S. Heifets, A. Wagner, B. Zotter, SLAC/AP110, January 1998
- [19] A. Burov, V. Danilov, PRL 82 11 15 March 1999
- [20] N. Mouquet, E. Métral, IPAC10 p2039
- [21] F. Roncarolo et al, PRST AB 12, 084401 (2009)
- [22] H. Tsutsui,CERN-SL-Note-2002-034 AP (2002)
- [23] E. Métral et al, EPAC06, p2919 or CERN-AB-2006-051.
- [24] B. Salvant et al, IPAC10, p2054
- [25] C. Zannini et al, IPAC'10 p2045

## 7.5.2 Beam-Based Characterization of Coupled Bunch Instabilities

*J.M. Byrd, LBNL*

Beam-based techniques for characterizing the coupled-bunch (CB) instabilities and the narrow-band impedances that drive them in a storage ring can be *passive* or *active*. Passive techniques apply when uncontrolled beam instabilities already exist in the ring and one can only characterize the frequency of the instabilities and attempt to identify the driving impedances. Active techniques are used when a feedback system or some other means to control the instabilities exists, allowing a measurement of the effect of the driving impedance (e.g. the growth rate).

### 7.5.2.1 Passive techniques

#### Multibunch beam spectrum [1]

Some of the characteristics of the driving impedances destabilizing a multibunch beam can be inferred from the spectrum of beam oscillation frequencies as observed on either a spectrum analyzer or an FFT of the digitized bunch signal. For a symmetric fill pattern of  $n_B$  bunches, the signal from a pickup in either sum or difference mode has frequency components at multiples of the bunch frequency,  $n_B f_0$ , with an amplitude proportional to the total beam current. The oscillations of individual CB modes appear as modulation sidebands about the bunch harmonics at a frequency of  $p n_B f_0 \pm (\ell f_0 + f_b)$ , where  $\ell$  is the CB mode index which ranges from 0 to  $n_B - 1$ ,  $p$  is an integer, and  $f_b$  is either the synchrotron or betatron frequency. In the longitudinal case, the sidebands are a combination of phase and amplitude modulation (PM and AM), depending on the dispersion function at the pickup location and the transverse sidebands are results of AM.

All  $n_B$  CB modes can be measured in a frequency span of  $p n_B f_0$  to  $p n_B f_0 + n_B f_0 / 2$  or from  $p n_B f_0 + n_B f_0 / 2$  to  $(p + 1) n_B f_0$ . In the case of a bunch in every rf bucket, all CB modes appear in each of the frequency spans  $0 - f_{rf} / 2$ ,  $f_{rf} / 2 - f_{rf}$ , etc. Shown in Fig.1 are some representative beam spectra showing unstable longitudinal CB oscillations for the case of 328 bunches filling all rf buckets. The center peak corresponds to an orbit harmonic. Except for the main bunch harmonic, these orbit harmonics result from an unavoidable slight asymmetry in the bunch charges. The sidebands are first order PM sidebands spaced by  $f_s$

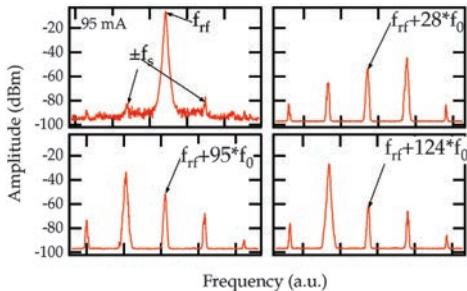


Figure 1: Raw frequency spectra near selected revolution harmonics showing unstable longitudinal CB oscillations.

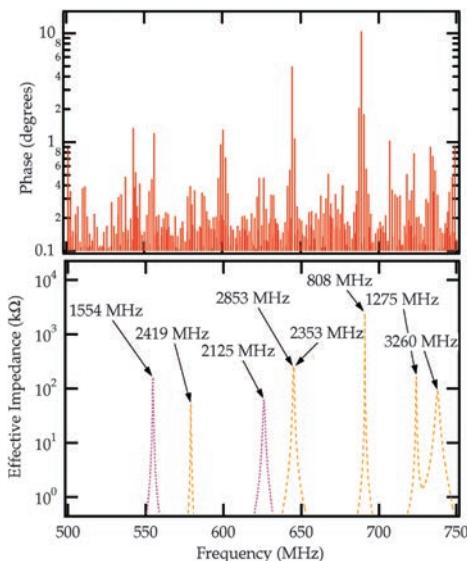


Figure 2: Measured spectrum of first-order longitudinal sidebands for 328 bunches in the frequency range  $f_{rf} - 3f_{rf}/2$ . The real part of the rf cavity impedance aliased into the same frequency range is shown below.

from the revolution harmonics, representing the amplitudes of different CB modes. The sidebands at  $2f_s$  are second-order PM sidebands.

Fig.2 shows the spectrum of all longitudinal CB modes measured in a bandwidth from 500 to 750 MHz. The patterns in the mode spectrum indicate the presence of resonant HOMs as the driving impedance. Identification of the HOMs is easier by comparing directly with the effective impedance ( $Z_{||}$ )<sub>eff</sub> (Sec.3.2.5), as determined from

either computation or bench measurement. The effective impedance can be considered as the actual HOM impedance aliased into the bandwidth determined by the bunch spacing. By measuring the CB mode spectrum at different symmetric fill patterns, the driving impedances can be unambiguously identified. The relative strength and bandwidth of the driving impedances can be inferred from the amplitude and width of the CB mode spectrum. However, it is difficult to achieve quantitative measurements directly using passive techniques except in the case where the HOM  $Q$  is large. In this case, the HOM  $Q$  can be inferred by measuring the affected CB mode as the HOM frequency is swept through the CB mode frequency. For example, consider the high- $Q$  mode at 808 MHz shown in Fig.2. The  $Q$  of this mode for each of two cavities was determined by measuring the CB mode amplitude as a function of the HOM frequency, which was tuned by slightly changing the cavity dimensions, e.g. by changing the cavity operating temperature.

**Single bunch cavity spectrum** It is often difficult to directly measure the spectrum of HOMs in an rf cavity once it is installed in the storage ring due to the restricted access to the cavity ports. However, a single-bunch beam can be used as an excitation mechanism and some of the properties of the HOMs can be measured from the spectrum of the cavity response. Because the spectrum of current excitation is a comb of lines at the revolution harmonics, this technique is most appropriate when the bandwidth of the HOM is at least  $f_0$ . Thus, this technique is more effective in larger rings or for cavities with moderately damped HOMs. Shown in Fig.3 is an example of the spectrum measured from a cell of an ALS third harmonic cavity. [2] The frequency and  $Q$  are fitted and the shunt impedance is found from the computed  $R/Q$ . This technique becomes difficult at frequencies approaching the beam pipe cutoff frequency because of the abundance of HOMs, some of which have no influence on the beam.

### 7.5.2.2 Active techniques

**Beam transfer functions** The impedance can be considered part of a feedback loop which includes both the beam and external excitation mechanism such as a feedback system. By measuring an open loop transfer function of the system, the complex impedance can be determined if

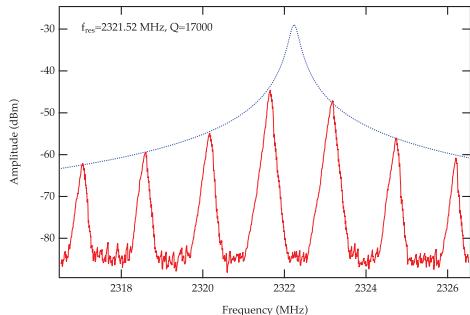


Figure 3: Single bunch spectra of a cell of the ALS third harmonic cavity showing a HOM near 2.32 GHz.

the frequency response of the excitation is known. This can be measured in the frequency domain with a swept frequency or white noise excitation or in the time domain with an impulse excitation. Note that the open loop transfer function is appropriate only below the instability threshold.

**Modulated feedback** If the CB instabilities can be controlled, e.g. by using a FB, a common technique for measuring the growth rate is to modulate the gain of the feedback. Most often the FB is turned off momentarily during which time the unstable CB modes grow with a rate proportional to the driving impedance. The FB is then turned on and damps the oscillations if they have not grown to an excessive level. This is commonly referred to as a “grow/damp” measurement. Both the driving impedance and the performance of the FB can be measured this way [3]. The growth of the unstable mode can be observed either on a spectrum analyser tuned to the mode or by making an FFT

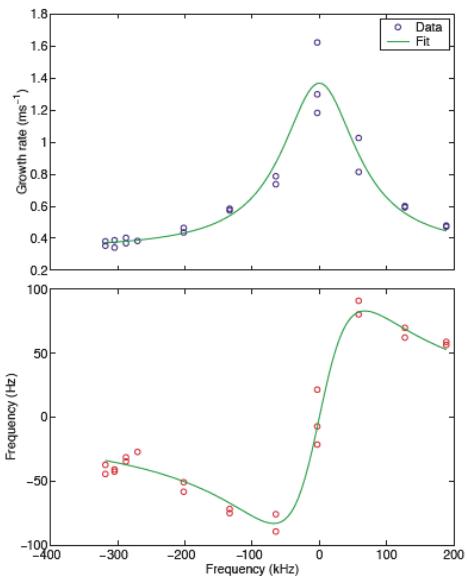


Figure 5: Complex frequency shift of a coupled bunch mode in the ALS as a HOM is tuned through the mode. The HOM impedance and  $Q$  can be determined from a fit to the data.

of the digitized bunch signals. Fig.4 shows an example of such an FFT measurement made at the KEKB LER [4]. The growth rate and frequency shift of the growing modes can be deduced from the grow/damp measurements, allowing determination of the effective driving impedance. An example of this is shown in Fig.5 [5] for the case of a HOM in the ALS RF cavity as the cavity HOM was tuned through the coupled bunch mode.

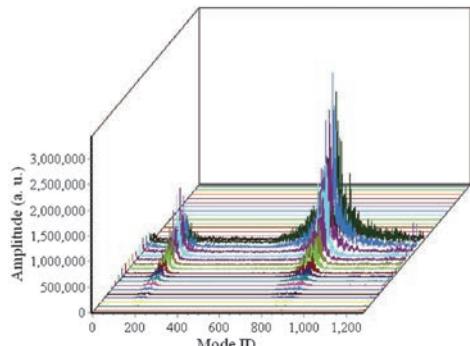


Figure 4: Growth of horizontal CB modes in the KEKB LER with the FB turned off.

## References

- [1] J.M. Byrd, J. Corlett, PA 51, 29 (1995)
- [2] J.M. Byrd et al, NIM A 455, 2 (2000)
- [3] R. Claus et al, EPAC96, WE005A, p.346
- [4] M. Tobiayama et al, PRST AB 9, 012801 (2006)
- [5] D. Teytelman et al, PRST AB 4, 112801 (2001)

### 7.5.3 Other Beam Based Methods to Measure Impedances

*J. Gareyte, CERN*

*J. S. Berg, BNL*

**Variation of transverse phase advance with current [1]** The variation with current of the

Table 1: Integral in Eq.(1) for various types of impedance. The “Multiplier” column must be multiplied by the value in the “Gaussian” or “Parabolic-like” column to get the value of the integral. For a resonator where  $\omega_R \sigma_z / \beta c \gg 1$ , the resonator impedance should be considered an inductive impedance of  $-iR_{\perp}/Q$ . The resonator impedance has value  $R_{\perp}$  at  $\omega = \omega_R$ .  $R_{\text{RW}}^{\perp} = (s_1 - s_0)/(2\pi b^3) \sqrt{Z_0 C/\pi \sigma_z}$ .

| Impedance   | Multiplier  | Gaussian      | Parabolic-like  |
|---|---|---------------|---|
| Inductive   | Imaginary part of impedance                             | $\sqrt{\pi}$  | $\frac{2^{2\mu-3}}{\sqrt{\mu}} \frac{\Gamma^2(\mu)\Gamma(\mu-1)}{\Gamma(\mu-1/2)\Gamma(2\mu-3/2)}$    |
| Resonator,<br>$\omega_R \sigma_z / \beta c \ll 1$ | $-(R_{\perp}/Q)(\omega_R \sigma_z / \beta c)^2$         | $2\sqrt{\pi}$ | $\frac{2^{2\mu-1/2}}{\Gamma(\mu+1/2)\Gamma(2\mu-1/2)} \sqrt{\mu} \Gamma^3(\mu)$                       |
| Resistive Wall                                    | $-R_{\text{RW}}^{\perp} \sqrt{2\pi \sigma_z / CT(1/4)}$ | 1             | $\frac{2^{2\mu-9/4}\Gamma^2(\mu)\Gamma(\mu-3/4)}{\sqrt{\pi}\mu^{1/4}\Gamma(\mu-1/4)\Gamma(2\mu-5/4)}$ |

Table 2:  $k_L$  for various impedances. The resonator impedance has value  $R_{\parallel}$  at  $\omega = \omega_R$ . The “Multiplier” column must be multiplied by the value in the “Gaussian” or “Parabolic-like” column to get the value of the integral.  $R_{\text{RW}}^{\parallel} = R_{\text{RW}}^{\perp} \pi b^2 / C$ .

| Impedance   | Multiplier   | Gaussian       | Parabolic-like   |
|---|--|----------------|--|
| Resonator,<br>$\omega_R \sigma_z / \beta c \ll 1$ | $R_{\parallel} \omega_R / 2Q$  | 1              | 1  |
| Resonator,<br>$\omega_R \sigma_z / \beta c \gg 1$ | $(R_{\parallel} \omega_R / Q^2)(\omega_R \sigma_z / \beta c)^{-3}$             | $\sqrt{\pi}/2$ | $\frac{\mu^{-3/2} 2^{2\mu-9/2}\Gamma^2(\mu)\Gamma(\mu-2)}{\Gamma(\mu-3/2)\Gamma(2\mu-5/2)}$            |
| Resistive wall                                    | $R_{\text{RW}}^{\parallel} \beta c / C (2\pi \sigma_z / C)^{-3/2} \Gamma(3/4)$ | 1              | $\frac{2^{2\mu-11/4}\Gamma^2(\mu)\Gamma(\mu-5/4)}{\sqrt{\pi}\mu^{3/4}\Gamma(\mu-3/4)\Gamma(2\mu-7/4)}$ |

betatron phase advance between two points in the ring can be used to estimate the imaginary part of the transverse impedance located between those two points.

$$\sum_k \beta_{\perp k} \int \text{Im} \left\{ Z_{\perp k} \left( \frac{x \beta c}{\sigma_z} \right) \right\} \left| \tilde{\rho} \left( \frac{x \beta c}{\sigma_z} \right) \right|^2 dx \\ \approx 4\pi \beta^2 \frac{\sigma_z}{C} \frac{E_0}{q} \frac{d(\Delta\psi_{\perp})}{di_B} \quad (1)$$

Here  $\Delta\psi_{\perp}$  is the betatron phase advance from longitudinal positions  $s_0$  to  $s_1$ , and  $\sum_k$  is a sum over all the impedance sources in that interval with transverse impedances  $Z_{\perp k}$ . Tab.1 gives the integral on the l.h.s. of Eq.(1) for various types of impedance.

In Fig.1, slope of the line in the arc sections gives impedance per unit length in those sections. For the sum of all arcs,  $d(\Delta\Psi_{\perp})/di_B = 2\pi \cdot 65 \text{ A}^{-1}$ , making the r.h.s. of Eq.(1)  $69 \text{ M}\Omega$ . Compare with l.h.s. of Eq.(1) of  $60 \text{ M}\Omega$  for the 2800 bellows alone ( $R_{\perp} = 0.4 \text{ M}\Omega/\text{m}$  at  $\omega_R = 2\pi \cdot 120 \text{ GHz}$ ,  $Q = 1$ ,  $\beta_{\perp} = 84.9 \text{ m}$ ). The two jumps give a total  $d(\Delta\Psi_{\perp})/di_B = 2\pi \cdot 55 \text{ A}^{-1}$ , from which the impedance of the cavities is similarly estimated.

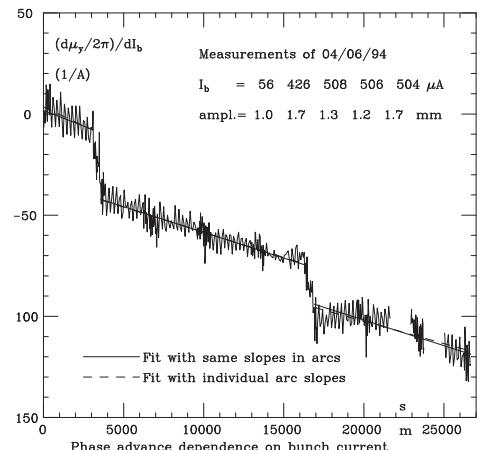


Figure 1: Derivative of transverse phase with current versus position in the ring, in LEP.

**Variation of orbit with current [1]** One can compute loss factor and thus get information about the real part of the longitudinal impedance based on shift of closed orbit with current. If  $Z_{\parallel}(\omega)$  is the total impedance between points  $s_0$

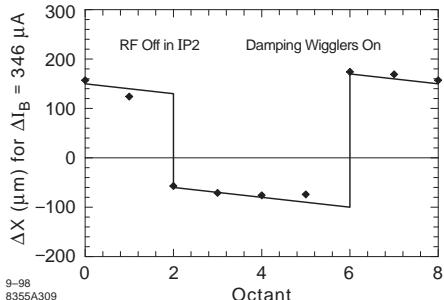


Figure 2: Closed orbit shift at a finite current versus position in the ring, in LEP.

and  $s_1$  in the ring, then

$$k_L = -\frac{(E_0/q)\omega_0}{2\pi} \frac{d}{di_B} \left( \frac{x_{co,1}}{D_{x,1}} - \frac{x_{co,0}}{D_{x,0}} \right) \quad (2)$$

$$k_L \equiv \frac{1}{2\pi} \int |\tilde{\rho}(\omega)|^2 \text{Re}\{Z_{||}(\omega)\} d\omega \quad (3)$$

Here  $x_{co,k}$  is the closed orbit displacement at  $s_k$ ,  $D_{x,k}$  is the dispersion at  $s_k$ . Tab.2 gives  $k_L$  for various types of impedance.

In Fig.2, slope of the lines in the arc sections gives the loss factor per unit length in those sections. The jump at point 2 (corresponding to  $d(x_{co,1} - x_{co,0})/di_B = 0.54$  m/A,  $D_x$  being the same) gives the loss in the rf cavities at that point (which are not powered).

### Variation of synchronous phase with current [2, 3]

$$\hat{V} \cos \phi_s = V_s + i_B k_L \frac{2\pi}{\omega_0} \quad (4)$$

where  $\hat{V}$  is the amplitude of the accelerating voltage,  $\phi_s$  is the phase of the bunch center counted from the crest of the voltage, and  $qV_s$  is the energy loss per turn due to synchrotron radiation.

## References

- [1] D. Brandt et al, PAC 95, p.570
- [2] M.A. Allen et al, PAC75, p.1838
- [3] W. Anders et al, XVth Int. Conf. High Energy Acc., (1992) p.1121

### 7.5.4 Direct Wakefield Measurement

W. Gai, ANL  
C. Jing, Euclid Techlabs, LLC

In fact, the direct wakfield measurement technique described here is a variation of the collinear

wakefield acceleration scheme. During the measurement, wakefields of the accelerating structure under test are excited by a short, high current drive bunch, which is followed by a low charge “witness” bunch. Depending on the temporal separation between two bunches, the witness bunch which experiences wakefields from the leading one can either gain or lose energy. While the delay time varies, energy change of the witness bunch is recorded in an energy spectrometer to profile either the short or long range wake potential of the drive bunch (wake potential can be easily converted to the wakefields). Both longitudinal and transverse wakes are amenable to this approach. The transverse wakefield is obtained by precisely creating an off-axis beam (the witness bunch keeps following the same position as the drive bunch at entrance of the accelerating structure). Energy difference of the witness bunch from the centered beam case reflects transverse wakefield of the structure.

A few key points for an accurate wakefield measurement include 1) using as intensive a drive bunch as possible to achieve a more measurable energy variation of the witness bunch; 2) using a short drive bunch to excite a broad frequency spectrum of the wakefield which is particularly important for the short range wakefield measurement; 3) using the witness bunch with a well defined energy separation from the drive bunch, which helps distinguish the witness from the drive beam in the energy spectrometer; 4) keeping the charge of witness bunch low and bunch length short to precisely probe wakefields from the drive bunch alone (note that the observed energy variation of the witness bunch results from the convolution of wakefields from both drive and witness bunches).

Witness bunch generation and its controllable delay line formation vary with the available resources. One example is the series of measurement performed at the Advanced Accelerator Test Facility (AATF) [1], where the 15-MeV witness bunch is generated by allowing 25% of the 20-MeV, 30-ps, 5-nC drive beam to strike a graphite target (see Fig.1). The intensity of the witness bunch is several orders of magnitude less than the drive beam, and it is transported in a different beam line before being recombined before the entrance of the testing structure. The adjustable witness bunch delay is formed by changing the physical length of the witness beamline, which

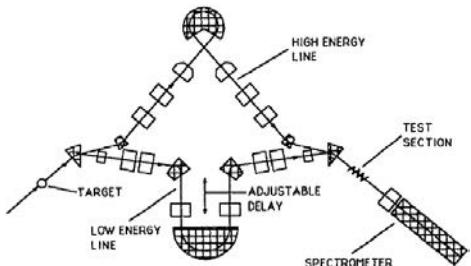


Figure 1: Plan view of the AATF.

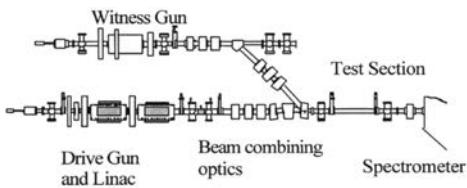


Figure 2: Schematic layout of the AWA facility.

is constructed with a section of vacuum bellow. The Argonne Wakefield Accelerator (AWA) facility, which is the replacement of the AATF, is another example of a platform to directly measure wakefields. As shown in Fig.2, the AWA facility has two L-band rf photoinjectors to provide independent drive and witness beams. Wakefields of accelerating structures in the test section can be mapped out by the witness bunch after it rejoins the drive beamline. The delay of the witness bunch is formed by the optical delay line of the laser beam to the witness gun. In this configuration, the totally uncoupled drive and witness bunches are able to hold the capability of the wakefield measurement in any time window.

As some examples of the achievable results, Fig.3 shows the measured and computed wakes for two dielectric-loaded accelerating structures. For details of the measurements refer to [2]. Currently the AWA drive beamline is being rebuilt to provide a 75 MeV high current drive beam ( $\sim$ GW level beam power).

In addition to directly measure energy change of the witness beam, experiments conducted at SLAC Accelerator Structure SETup (ASSET) provides another option to profile transverse wakefield using a witness beam [3]. The setup is shown in Fig.4, where a positron beam induced wakfields in the structure under test that

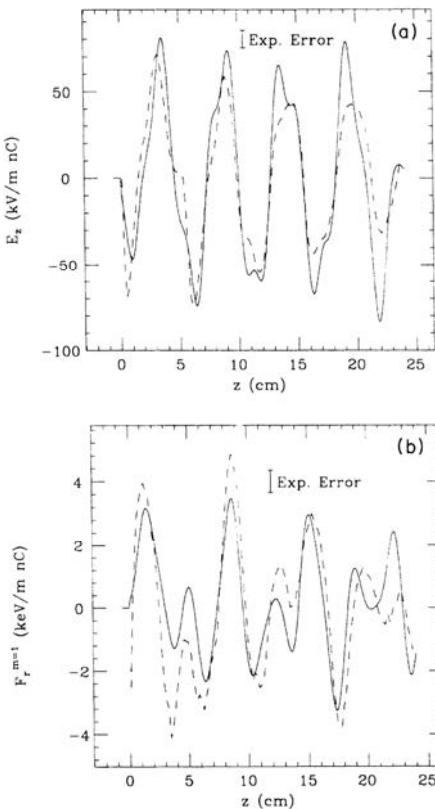


Figure 3: Comparison of experimental data with calculation of the longitudinal (a) and transverse (b) wakefields in a dielectric loaded accelerating structure [2]. The dashed line represents data and the solid line is the theory (first four modes).

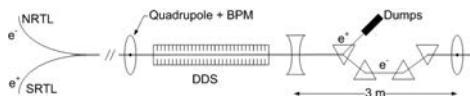


Figure 4: Layout of ASSET in the SLC [3].

are witnessed by an electron beam. Just downstream of the structure, a chicane of magnets steers the positrons to a dump and the electrons into the linac where their trajectory was recorded by Beam Position Monitors (BPMs). The transverse wakefield was determined by measuring the change of witness bunch angular trajectory due to deflection from the off-centered drive beam, which is proportional to dipole wake potential of

## Sec.7.6: POLARIMETERS

the measured structure for a certain drive beam offset.

### References

- [1] E. Chojnacki et al, JAP 69 (1991) 6257
- [2] M. Rosing, W. Gai, Phys. Rev. D 42 (1990) 1829
- [3] C. Adolphsen et al, SLAC-PUB-7519, 1997

## 7.6 POLARIMETERS

### 7.6.1 Lepton Polarimeters

*E. Chudakov, D. Gaskell, J. Grames,  
JLab, M. Woods, SLAC*

The polarization of an electron beam is typically measured by Mott (<10 MeV), Möller (>100 MeV) or Compton (>1 GeV) scattering polarimeters.

#### 7.6.1.1 Mott polarimetry

The polarized electron elastic differential cross section as a function of scattering angle  $\theta$  from the nuclear charge  $Z$  is

$$\sigma(\theta, \phi) = I(\theta)[1 + S(\theta)\vec{P} \cdot \hat{n}], \quad (1)$$

where  $I(\theta)$  is the unpolarized cross section

$$I(\theta) = \left( \frac{Ze^2}{2mc^2} \right)^2 \cdot \frac{(1 - \beta^2)(1 - \beta^2 \sin^2(\frac{\theta}{2}))}{\beta^4 \sin^4(\frac{\theta}{2})}, \quad (2)$$

$S(\theta)$  (Sherman function) is the analyzing power,  $\vec{P}$  is the electron polarization,  $\hat{n} = \frac{\vec{k} \times \vec{k}'}{|\vec{k} \times \vec{k}'|}$  is the unit vector normal to the scattering plane where  $\hbar\vec{k}$  ( $\hbar\vec{k}'$ ) is the incoming (outgoing) electron momentum. Eq.(1) implies the beam polarization normal to the scattering plane contributes a non-zero scattering asymmetry, so pairs of detectors are used to measure, e.g., the horizontal or vertical transverse polarization components [1].

The Sherman function [2] for a typical high-Z atom and electron energies are shown in Fig.1. Uncertainty of the Sherman function is dominated by electron cloud screening (few percent) < 1 MeV [3] or by finite nuclear size penetration [4] (1%) above 1 MeV. The measured asymmetry is diluted by plural and multiple scattering as target thickness increases. An effective analyzing power is determined by extrapolating asymmetries from various thickness targets to zero thickness and using the Sherman function for absolute calibration.

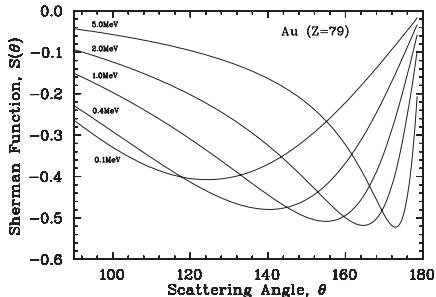


Figure 1: Sherman function vs. scattering angle.

Table 1: Mott polarimeters.

| Polarimeter      | Energy (MeV) | Target   | $(\frac{\partial P}{\partial \theta})_{sys}$ |
|------------------|--------------|----------|--|
| [5] TU Darmstadt | 0.1          | Au,Ag    | 3.0%   |
| [6] KPH Mainz    | 1-3.5        | Au       | 2.0%   |
| [7] JLab CEBAF   | 2-8          | Au,Ag,Cu | 1.1%   |

Mott polarimeters are listed in Table 1. Low energy (< 200 keV) Mott polarimeters are suited for polarized electron sources, but suffer from high cross section at large Sherman function ( $\theta < 125^\circ$ ) requiring nA current, even with thin (<100 nm), difficult to manage, target foils. Clever techniques are employed [8], yet plural and multiple scattering dilute the elastic asymmetry. High energy (> 1 MeV) Mott polarimeters are suited for accelerators; allowing good resolution scintillators and time-of-flight or spectrometer discrimination to isolate the elastic signal; lead shielding and veto detectors suppress the photon background. Plural scattering is significantly reduced and target thickness extrapolation can be precisely modeled [7]. Low cross section allows  $\mu\text{A}$  current, manageable  $\mu\text{m}$  thick targets, and beam position/intensity monitors to control systematics. Elastic rates from gold at the maximum Sherman function are 0.1-1 kHz/ $\mu\text{A}/\mu\text{m}$ .

#### 7.6.1.2 Möller polarimetry

This makes use of polarization asymmetries in the  $e^-e^-$  scattering cross section [9]:

$$\frac{d\sigma_M}{d\Omega} = \frac{\alpha^2}{s} \frac{(3 + \cos^2 \theta')^2}{\sin^4 \theta'} \left[ 1 - P_{\parallel}^b P_{\parallel}^t A_{\parallel}(\theta') - P_{\perp}^b P_{\perp}^t A_{\perp}(\theta') \cos(2\phi' - \phi'_1 - \phi'_2) \right] \quad (3)$$

where  $\theta', \phi' =$  CM electron scattering angles,  $\phi'_1, \phi'_2 =$  azimuths of transverse polarization vectors,  $P_{\perp, \parallel}^{b,t} =$  beam and target transverse/longitudinal polarization states.

The analyzing power is defined by the transverse and longitudinal asymmetry functions  $A_{\perp,\parallel}$  (max for 90° scattering in CM):

$$A_{\perp} = \frac{\sin^4 \theta'}{(3 + \cos^2 \theta')^2} \leq \frac{1}{9} \quad (4)$$

$$A_{\parallel} = \frac{(7 + \cos^2 \theta') \sin^2 \theta'}{(3 + \cos^2 \theta')^2} \leq \frac{7}{9} \quad (5)$$

The analyzing power does not depend on the beam energy and only weakly depends on the scattering angle around 90° in the CM frame. Møller polarimeters select secondary particles with energies close to half the beam energy and produced in a certain angular range. Various magnetic optics schemes are used (see Table 2). Detecting both scattered electrons allows to reduce the non-Møller backgrounds to negligible levels.

The polarized electron targets consist of thin ferromagnetic foils magnetized in external magnetic fields. The average polarization of electrons in saturated iron at room temperature is ~8.04%[14]. In the typical scheme the foil oriented at a small angle (~ 20°) to the beam is magnetized along its surface by a field of 0.01–0.03 T parallel to the beam. Typically, only the longitudinal beam polarization is measured, but a transverse component of the target polarization allows to measure the transverse beam polarization as well. In a different scheme the foil is oriented perpendicular to the beam and is magnetized to saturation by a 3-4 T field perpendicular to its surface[14]. The latter method reduces the target polarization uncertainty from 2-3% to sub-percent level[12].

The spectrometer acceptance may depend on the initial momentum of the target electron (comparable to  $e^{\pm}$  rest mass for inner shells of iron atoms), which changes the effective target polarization and requires a correction [15].

Fig.2 shows the SLAC E-154 Møller Polarimeter [10]. Parameters of several polarimeters are summarized in Table 2.

Table 2: Møller polarimeters.

| Polarimeter  | Arms | Optics | $(\frac{\partial P}{\partial P})_{sys}$ |      |
|--------------|------|--------|---|------|
|              |      |        | target                                  | full |
| [10] SLAC    | 1    | D      | 1.7%                                    | 2.7% |
| [11] Bates   | 2    | Q      | 1.5%                                    | 2.9% |
| [13] JLab, A | 2    | QQD    | 2.0%                                    | 3.0% |
| [12] JLab, C | 2    | QQ     | 0.3%                                    | 0.5% |

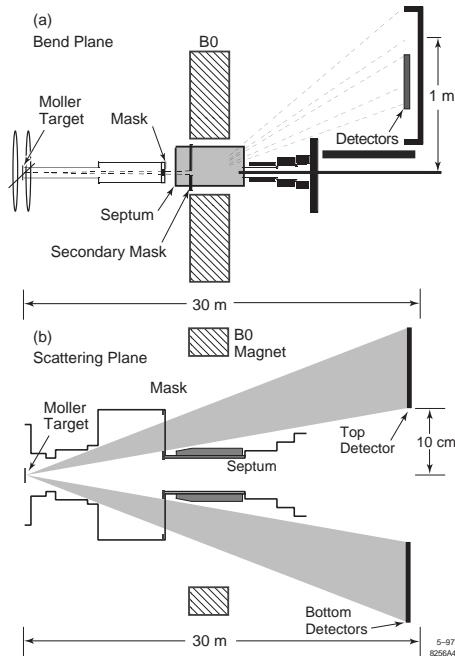


Figure 2: The SLAC Møller Polarimeter[10].

### 7.6.1.3 Compton polarimetry

Compton scattering can be used to measure both longitudinal and transverse components of the beam polarization. Useful references and formulae can be found in [16, 17]. Here we adopt the formalism given in [16]. The unpolarized differential Compton cross section can be expressed as,

$$\left( \frac{d^2\sigma}{dxd\phi} \right)_{unp} = r_0^2 y \left\{ \frac{x^2(1-y)^2}{1-x(1-y)} + 1 + \left[ \frac{1-x(1+y)}{1-x(1-y)} \right]^2 \right\} \quad (6)$$

where  $r_0$  is classical electron radius and the dimensionless  $x$ ,  $y$ -scattering parameters are defined in terms of the incident (scattered) electron and photon energies,  $E_0$  and  $\omega_0$  ( $E$  and  $\omega$ ),

$$y = \left( 1 + \frac{4E_0\omega_0}{m_e^2} \right)^{-1} \quad x = \frac{\omega}{\omega_{max}} \quad (7)$$

with  $\omega_{max} = E_0(1-y)$ .

The polarized cross section is then,

$$\frac{d^2\sigma}{dxd\phi} = \sigma_{unp}(1 - P^{\gamma}[P_z^e A_z(x, y) + P_t^e \cos \phi A_t(x, y)]), \quad (8)$$

where  $\sigma_{unp} = \left( \frac{d^2\sigma}{dxd\phi} \right)_{unp}$ ,  $\phi$  is the azimuthal angle of the outgoing photon with respect to the electron transverse polarization,  $P_\gamma$  is the incident photon's circular polarization,  $P_z^e$  and  $P_t^e$  are the incident electron's longitudinal and transverse polarization, and the longitudinal and transverse analyzing powers are given by,

$$A_z(x, y) = \frac{r_0^2 y}{\sigma_{unp}} [1 - x(1+y)] \times \{1 - [1 - x(1-y)]^{-2}\} \quad (9)$$

$$A_t(x, y) = \frac{r_0^2 y}{\sigma_{unp}} x(1-y) \frac{\sqrt{4xy(1-x)}}{1-x(1-y)}. \quad (10)$$

$P_z^e$  ( $P_t^e$ ) can be extracted via Equations 8 and 9 (8 and 10).  $A_z$  is maximal at the kinematic endpoint  $E_{min} = yE_0 = E_0 + \omega_0 - \omega_{max}$ , corresponding to  $180^\circ$  backscattering in the center-of-mass (COM) frame; it is zero for  $90^\circ$  scattering in the COM.

Figure 3 shows an example of the Compton cross section and longitudinal and transverse analyzing powers vs. the scattered photon energy. Table 3 lists results for several Compton polarimeters that measure  $P_z^e$ . The HERA polarimeter measuring  $P_t^e$  is described in [21].

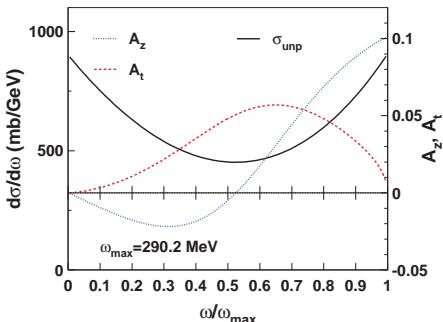


Figure 3: Compton cross section and analyzing powers versus  $(\omega/\omega_{max})$  for a 532 nm laser and a 3 GeV electron beam.

Table 3: Compton polarimeters.

| Polarimeter    | Energy (GeV) | $(\frac{\sigma P}{P})_{sys}$ |
|----------------|--------------|------------------------------|
| [18] NIKHEF    | <1           | 4.5%                         |
| [19] JLab, A   | 1-6          | 1.0%                         |
| [20] HERA LPOL | 27.5         | 1.6%                         |
| [22] SLD SLAC  | 45.6         | 0.5%                         |

## References

- [1] J. Kessler, Polarized Electrons 2nd Ed., Springer-Verlag (1985)
- [2] N. Sherman, PR 103 (1956) 1601
- [3] S. Lin et al, Nucl. Phys. 45 (1963) 492
- [4] P. Ugincius et al, Nucl. Phys. A158 (1970) 418
- [5] R. Barday et al, PST 2009 Proceedings (2010)
- [6] V. Tioukine et al, PST 2009 Proceedings (2010)
- [7] M. Steigerwald, AIP 570 (2001) 935
- [8] T.J. Gay, F.B. Dunning, RSI 63 (1992) 1635
- [9] C. Møller, Ann. Phys. (Leipzig) 14 (1932) 532; A. A. Kresnin, L. N. Rosentsveig, Soviet JETP 5 (1957) 288
- [10] H.R. Band et al, NIM A 400 (1997) 24
- [11] J. Arrington et al, NIM A311 (1992) 39; K. B. Beard et al, NIM A361 (1995) 46
- [12] M. Hauger et al, NIM A462 (2001) 382
- [13] A. V. Glazulin et al, Fizika B 8 (1999) 91 [arXiv:hep-ex/9912063].
- [14] L.V. de Beveret al, NIM A 400 (1997) 379
- [15] L.G. Levchuk, NIM A 345(1994) 496
- [16] M.L. Swartz, SLAC-PUB-4656 (1988)
- [17] I.F. Ginzburg et al NIM A 219 (1984) 5
- [18] I. Passchier et al, NIM A 414 (1998) 446
- [19] M. Baylarc et al, PL B 539 (2002) 8; A. Acha et al , PRL 98 (2007) 032301
- [20] M. Beckmann et al, NIM A 479 (2002) 334
- [21] D.P. Barber et al, NIM A 338 (1994) 166; B. Sobloher, PST 2009 Proceedings (2010)
- [22] M. Woods, HE Spin Physics 1996:843; K. Abe et al, PRL 84 (2000) 5945

## 7.6.2 Proton Polarimeters

*Y. Makdisi, BNL*

The degree of beam polarization along a certain direction is defined as:

$$P = \frac{n^+ - n^-}{n^+ + n^-}$$

Where  $n^+$  and  $n^-$  are the number of protons with spins parallel and antiparallel to the stable spin direction. In an accelerator, the stable spin direction is usually transverse to the momentum vector and along the vertical. Beam polarization is typically measured using a nuclear reaction in a plane perpendicular to the polarization direction. In addition to static beam polarization measurements, polarimeters are also used during the acceleration ramp to detect polarization loss due to resonances. An accuracy of 5% or better represents the current norm for high energy proton beam polarimeters.

A single arm spectrometer measures the number of events detected with the beam polarized up

Table I

| Energy            | Reaction  | Kinematic Region                          | Analyzing Power                              |
|-------------------|---|---|--|
| < 0.1 GeV [1, 2]  | $p + {}^{12}\text{C} \rightarrow p + {}^{12}\text{C}$ | $\theta \approx 50^\circ$                 | $A \approx 0.8 \dots 0.98$                   |
| 0.1 ... 1 GeV [3] | $p + {}^{12}\text{C} \rightarrow p + X$               | $\theta \approx 16^\circ$                 | $A \approx 0.2 \dots 0.6$                    |
| 1 .. 20 GeV [4]   | $p + p \rightarrow p + p$                             | $t \approx -0.15 (\text{GeV}/c)^2$        | $A \approx 0.7/P_{\text{lab}}(\text{GeV}/c)$ |
| > 20 GeV [5, 6]   | $p + {}^{12}\text{C} \rightarrow \pi^{+(-)} + X$      | $Pt \approx 0.5 \text{ GeV}/c, X_F = 0.5$ | $A \approx 0.15$                             |
| > 20 GeV [7]      | $p + p \rightarrow p + p$                             | $t \approx -0.002 (\text{GeV}/c)^2$       | $A \approx 0.04$                             |
| > 20 GeV [8]      | $p + {}^{12}\text{C} \rightarrow p + {}^{12}\text{C}$ | $t \approx -0.002 (\text{GeV}/c)^2$       | $A \approx 0.03$                             |
| 62-250 GeV [9]    | $p + p \rightarrow n + X$                             | $\theta \approx <3 \text{ mrad}$          | $A \approx 0.09$                             |
| 100 GeV [10]      | $p + p \rightarrow \text{charged incl.}$              | $0.38 < \theta < 1.15 \text{ degrees}$    | $A \approx 0.007$                            |

versus those with the beam polarized down taking into account the proper normalizations and that the experimental conditions remain unchanged. The resulting beam polarization is:

$$P = \frac{1}{A} \frac{N^+ - N^-}{N^+ + N^-}$$

$N^+$  and  $N^-$  are the number of scatters, normalized to the incident beam, with the beam polarization up and down respectively and  $A$  is the analyzing power of the reaction. Similarly, a two arm apparatus that measures the number of scatters to beam-left,  $N_L$ , and beam-right,  $N_R$ , simultaneously will determine the degree of up and down beam polarization independently and one substitutes  $N_L$  and  $N_R$  in the above equation. The associated statistical error in these measurements is a function of both the analyzing power  $A$  and the total number of events  $N$  [ $(N^+ + N^-)$  or  $(N_L + N_R)$ ]

$$\Delta P = \frac{1}{A} \frac{1}{\sqrt{N}}$$

In general both methods are utilized and combined when the beam polarization is alternated between the two states (along and opposite the stable spin direction), in order to reduce sensitivity to systematic errors and potential dependence on geometrical effects.

$$P = \frac{1}{A} \frac{\sqrt{N_L^+ N_R^-} - \sqrt{N_L^- N_R^+}}{\sqrt{N_L^+ N_R^-} + \sqrt{N_L^- N_R^+}}$$

and

$$\Delta P = \frac{1}{A} \frac{1}{\sqrt{N_L^+ + N_R^+ + N_L^- + N_R^-}}$$

It is therefore important to utilize reactions with large analyzing power as well as large cross section. In the design of a polarimeter, the figure of merit is the product ( $N \cdot A^2$ ).

Unlike electromagnetic reactions (see the previous section) the analyzing power and cross section of proton induced nuclear reactions are not precisely calculable especially at high energies, thus one reverts to experimental results. These measurements are generally done with polarized proton targets the polarization of which is well measured using NMR techniques and Masers. As these targets are not pure hydrogen, the target material presents undesired background especially in inclusive measurements. However, this is less of an impediment for exclusive reactions such as p-p elastic scattering when both outgoing particles are measured.

At relatively low energies below 12 GeV good accuracy 3%-5% can be attained in beam polarization measurements primarily using p-p elastic scattering at the ANL ZGS, Saturne at Saclay and COSY at Jülich. The analyzing power of this reaction scales as  $1/P_{\text{Lab}}$  and becomes increasingly small resulting in a reduced figure of merit and accuracy.

At Brookhaven National Laboratory, polarized beams span the range from the 200 MeV LINAC through the AGS (3-24 GeV) and then RHIC (24-250 GeV). Several polarimeter concepts were developed to cover the RHIC energy range. These included inclusive pion production at large  $X_F$  (requiring significant spectrometers) and p-carbon elastic scattering in the Coulomb Nuclear Interference (CNI) region (compact spectrometers, 2% accuracy in less than 1 minute). In addition, a polarized hydrogen jet target was installed to measure the beam polarization and calibrate the polarimeters at any desired energy using pp elastic scattering in the CNI region [7] (a 5% statistical accuracy per 6 hour fill) and calibrated the RHIC polarimeters to the desired 5% accuracy at top RHIC energies. At RHIC it was

also discovered that inclusive neutron production in the extreme forward direction carried a significant analyzing power that rendered it useful as a local polarimeter utilized by the PHENIX and STAR experiments. The latter also employs an additional local Beam-Beam Counters (BBC) polarimeter that measures inclusive charged particle production at moderate angles to the beam. Table I provides a list of the various reactions employed and their respective analyzing powers.

## References

- [1] W. Haeberli et al, NIM 163 (1979) 403
- [2] S. Kato et al, NIM 169 (1980) 589
- [3] M.W. McNaughton et al, NIM A241 (1985) 435
- [4] H. Spinka et al, NIM 211 (1983) 239
- [5] D.L. Adams et al, PL B264 (1991) 462
- [6] C.E. Allgower et al, PR D65 (2002) 092008
- [7] H. Okada et al, PL B638 (2006) 450; I.G. Alekseev et al, PR D79 (2009) 094014
- [8] I. Nakagawa et al, Eur. Phys. J. ST 162 (2008) 259-265
- [9] Y. Fukao et al, PL B650 (2007) 325
- [10] C.A. Whitten Jr., AIP Conf. Proc. 980, PSTP 2007, pp. 390-396

## 7.7 CONTROLS AND TIMING

*K. Rehlich, DESY*

The complexity of control systems is constantly increasing. Recent accelerators implement 500 crate systems (e.g. VME or similar). The design of a control system needs careful planning and strategies to keep the system simple and well structured with loose coupling and adequate interfaces between subsystems and modules. A further challenge is the rapid change in computer hardware and software. Since software development has to start early to be ready in time, a careful decision has to be taken with a balance between a mature and reliable technology and, on the other hand, modern technologies that are not outdated when the accelerator is commissioned. Descriptions of some control systems are available on the web [1]-[6], most of them can even be downloaded. Very often control systems are not uniform designs; different new technologies and interfaces are added during their lifetime or by contributions from collaboration partners. This mixture is in general no problem as long as all subsystems can be integrated by well-defined interfaces.

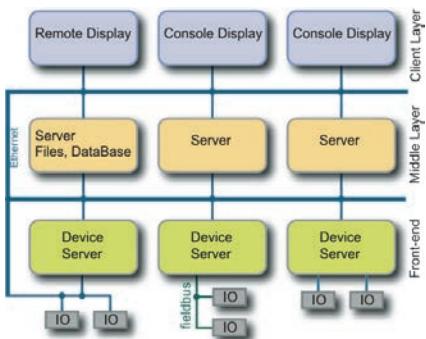


Figure 1: Overall architecture.

The evolution in the computer industry and new future accelerators like the ILC (International Linear Collider) with much higher data volumes will keep control systems development dynamic.

**Architecture** As shown in Fig.1, the overall architecture consists of three layers of computers. This is called often the ‘Standard model’ of control systems. A top level with display or client programs as a presentation layer, a front-end level with device servers and I/O and a middle layer with powerful data servers. The upper services with the application programs are available to the consoles in the control room, the experts working at the machine or in their offices. For remote operations all these applications are provided via secure links to the remote shifts or experts. Examples of applications in the presentation layer are: archive viewer, synoptic displays with a graphic representation of the subsystems, alarm panel, save and restore tool, plotting and measurement and simulation tools. Client programs often require information from a group of devices or need to operate on such a group. Middle layer servers can provide such collective reports and controls. An example of an implementation is a Finite State Machine (FSM) server. This server keeps track of several devices and is able to initiate state dependent control actions. A hierarchical grouping of servers is a way to better structure the system. These kinds of servers are usually implemented in the middle layer. Data bases, file servers, machine physics simulation servers, name servers, slow feedbacks and other automation as well as Web services, and in general services without a direct device connection, are placed in the middle layer.

The front-ends are distributed and placed close to the devices. In former years the institutes developed a lot of front-end electronics. Nowadays most of the electronics is based on industrial standards and housed in crate systems like VME [7], Compact PCI (cPCI) or in Personal Computers with a PCI/PCIe bus. Recently the new systems ATCA and  $\mu$ TCA [8] were introduced for the telecom industry. The design is made for very high availability and allows life-insertion of modules with the required software handling. A full management on crate and module level is included in the standard. For applications in physics, several institutes and companies formed a working group within PICMG to add features to the standard for accelerator applications. This standard is called MTCA.4 "Micro TCA enhancements for Rear I/O and Precision Timing". All bus standards define additional mezzanine cards that can be plugged on CPU boards or special carrier boards. Examples are PMC (with a PCI bus), IP (Industrial Pack) and AMC for the new TCA standards. AMC has in addition the feature to be software compatible with ATCA and can be used stand-alone in a  $\mu$ TCA crate. The mentioned crate systems are used for fast acquisitions and feedbacks. Slow and reliable tasks are better implemented in Programmable Logic Controllers (PLC). Interlocks or the machine protection system are typical applications for PLCs. The control system is connected to the PLC by a fieldbus, like CAN or ProfiBus, or an Ethernet link.

A fieldbus is a network that connects sensors, actuators or complex front-end input/output devices to the local device servers. Ethernet seems to replace the fieldbuses more and more. Stepper motor controller, temperature sensors, digital I/O and ADCs are examples of other devices connected with fieldbuses to the control system. Recently complete "supervisory control and data acquisition" (SCADA) systems are integrated in control systems. SCADA uses field buses together with the software to configure and operate a factory plant. The LHC in CERN uses PVSS for example.

In future the parallel bus systems will be replaced by serial point-to-point Gigabit links. Bus systems for high speed suffer from the insufficient RF properties of many connectors and stub traces on the boards. Therefore PCI will be extended with PCI-Express and VME by VXS. In these new standards the boards are connected by two or more Gigabit serial links via a star fabric

as a switch and router for the data traffic. A fabric allows several boards communicating at the same time. ATCA/ $\mu$ TCA is only based on serial links like Gigabit Ethernet, Serial Rapid IO or PCI-Express. FPGAs (Field Programmable Gate Arrays) are dominating the hardware market. Recent versions implement one or more processor cores and Gigabit serial links. It was demonstrated that a complete front-end server could run on a FPGA.

The communication protocol together with the API (Application Program Interface) is the glue of the control system. TCP or UDP are the protocols used on top of the Ethernet hardware.

Recent control systems are using a further protocol layer on top of UDP or TCP: CORBA (Common Object Request Broker Architecture), RPC (Remote Procedure Calls), DDS (Data Distribution Service) and JMS (Java Message Service) are the main examples. Most common services provided by the API are synchronous or asynchronous get or set functions, subscriptions and notifications to receive instantaneous updates from servers and broadcast services. Fast feedback systems require low latency links without protocol overhead and are often implemented in raw point-2-point Giga links between FPGAs.

**Network** In modern control systems the front-ends are connected by fieldbuses or Ethernet. Digital cameras are attached nowadays by USB, Ethernet or FireWire. All the rest communicates by Ethernet via switches and routers. The used Ethernet speeds in the field are 100, 1000 Mbit/second which results in a maximum data rate of about 10 or 100 MB per second respectively. The backbone of the Switches can be connected with higher speeds. Recent standards define 40 and 100 Gbit/second. Usually the control system is decoupled from the outside Internet by one or more firewalls to secure the control system network from attacks or viruses. Network security management is of growing importance. It has to balance the user requirements of a comfortable access to the control system from various places and the restrictions imposed by the increasing number of attacks and other unpredictable service degradations. Measures have to be taken to monitor activities and to apply access restrictions proactively or in case of an intrusion. A clean structure of the network is mandatory. Internal problems caused by e.g. hardware failures or configuration mistakes, even in single computers, might disturb the network as well. Isolation of slow industrial devices in separate subnets can

prevent them from corruption since network protocols are not always correctly implemented. Protocol features might be extended in future and misinterpreted in some devices. The tasks for the network management are further increased when remote operations or maintenance is considered.

On the other hand the good experience with existing remote operations shows that it is feasible without too much risk. A fast help from a remote expert at home can reduce the downtime considerably. The main goal of the network security should be to increase the availability of the investment.

Notebooks are equipped with wireless networks. Maintenance work on the electronics and front-end devices usually require access to the control system data. A Wi-Fi or WLAN (Wireless Local Area Network, e.g. 802.11 g or n) infrastructure has proven to be very useful and provides sufficient speed (about 54 Mbit/s, n is about a factor 6 faster) for maintenance tasks. But distortions of sensitive electronic equipment by wireless networks and mobile phones have to be considered. Radiation protection is a further concern if the equipment is installed in an accelerator building.

**Software** Control system software covers the whole range from server programs, communication protocols, applications and some Web services. With the increasing expectations of the users, the growing demands on availability of the accelerators, the need of more automation and on the other hand a growing complexity of the software products, the control system software becomes a critical issue. Operating systems on the servers have changed from real-time systems Like VxWorks and OS9 to LINUX and other UNIX flavors, on the client computers Windows, LINUX, and Mac OS are used most often. Also the used programming languages are drifting, from C to object-oriented designs, mainly based on the languages C++ and Java. Object-orientation is more than a hype - the idea is to get a better structured and modular architecture that is easier to maintain and improve. The growing complexity of the software development requires adequate tools: Integrated Development Environments (IDEs) for the complete design cycle from UML (Universal Modeling Language) models, editors, GUT (Graphical User Interface) designers, compilers, debuggers, performance analyzers, documentation generation and code repositories. A centralized repository like CVS (Concurrent Versions System) is mandatory for developers working in

a team or in collaborations with members in different institutes. A further tendency is the integration of commercial software. SCADA systems, MATLAB for mathematical processing or LabVIEW for data acquisition are a few common examples. Interesting is the development of using ‘open source’ products and even to publish control systems as ‘open source’ projects. Web technologies and services are used by control systems or are attached to them. XML (eXtensible Markup Language) is such a technology to describe and store all kind of complex data e.g. configurations. The rich capabilities of Web services are used to connect control system data with information repositories.

**Operation** Beside the “traditional” control system applications some further tools are required for the operation of an accelerator. Paper based logbooks are nowadays replaced by eLogBooks. This electronic version has a much better quality, allows searches of content and any application program can print into it. Remote operations certainly require eLogBooks. But the main advantage is the involvement of experts in the machine operation since one can see results and problems from any Web browser. From experience and from various discussions in Global Accelerator Network workshops (GAN) the feasibility of remote maintenance and even remote operation was demonstrated. The most practical approach is to run application programs on a computer at the accelerator site and to transfer the display information via the X11 (UNIX windowing system) or VNC (Virtual Network Computing) protocol to the remote site. In addition video conferencing systems are used for the communication during remote shifts and meetings. EVO is such a video-conference system that is available free of charge for several operating systems [9].

A control system demands manpower during the whole lifetime of an accelerator. It needs bug fixes, improvements, extensions, security patches and also hardware repairs. The operability and availability of the facility is to a substantial amount influenced by the control system. Automated operation, fast detection of problems and reasons of faults and good error recovery are the positive factors of software. But a control system might also cause a downtime. A Mean Time Between Failures (MTBF) of 67 hours was measured at Jefferson Lab for the software. With increasing complexity of the facilities and number of equipment the hardware failure rates are

becoming a serious issue. Some examples of MTBF numbers are: PC motherboard  $5 \times 10^4$  hours, VME CPU  $18 \times 10^4$ , Fan  $5 \times 10^4$ , SCSI disk  $10^6$ , IDE disk  $3 \times 10^5$ , power supply  $4 \times 10^4$  to  $10^5$ , and integrated circuit  $3 \times 10^7$  hours. These values vary considerably with temperature, radiation, power dissipation or operation close to limits e.g. voltage ratings. And these numbers scale down linearly with the number of devices. Ten required subsystems with a 99% uptime culminate in a 90% total availability or 100 PC motherboards will have a failure every three weeks in average. A redundant layout improves the availability, e.g. two independent systems operate in parallel.

$$\text{Availability} = \frac{\text{MTBF}}{\text{MTBF} * \text{downtime}}$$

Decreasing the downtime or time to repair by e.g. using hot-swap components and a good failure notification system, is a way to come closer to 100% availability. Anyway it is good practice to avoid ‘single points of failures’ and to couple multiple systems only loosely.

**Timing** The purpose of a timing system is to trigger devices on certain events of the machine operation, to provide synchronized clocks for ADC sampling and to distribute further reliable information. Typically distributed devices like the gun, kickers, klystrons, data acquisition modules etc., need to be triggered with configurable delays from a common source. Some of them require in addition a further synchronization for the ADC sampling. Triggers need drifts and jitters below one ns. ADC clocks might require a much better stability. Timing signals have to be stable with reliable delivery and are therefore implemented in hardware. The parameters of the timing system should be controlled and readable by the control system. Several designs with different requirements for the resolution and jitter parameters exist. The basic frequency of timing systems is in the range of 400 MHz to 2.7 GHz since common semiconductor technology supports this range. This frequency is derived from a RF source (master oscillator) of the accelerator. Pulsed machines are often synchronized to the mains (50 or 60 Hz) in addition to reject line frequency distortions of sensitive electronics. A central event generator distributes triggers, events and further information as encoded data on a fixed frequency. A single wire transports the signals from the event-generating module to the event-receiving units.

Tree, chained or redundant topologies allow a very flexible distribution of events. Event type and delay times can be programmed in the receiving modules to generate trigger signals, variable gates or different clock frequencies.

## References

- [1] ACS: [www.eso.org/gchiozzi/AlmaAcs](http://www.eso.org/gchiozzi/AlmaAcs)
- [2] Cern: <http://controls.web.cern.ch/Controls>
- [3] DOOCS: <http://doocs.desy.de>
- [4] EPICS: [www.aps.anl.gov/epics](http://www.aps.anl.gov/epics)
- [5] TANGO: [www.esrf.fr/Infrastructure/CComputing/tango](http://www.esrf.fr/Infrastructure/CComputing/tango)
- [6] TINE: <http://adweb.desy.de/mstltine/>
- [7] Vita (VME): [www.vita.com](http://www.vita.com)
- [8] μTCA, AMC, ATCA: <http://www.picmg.org/v2internal/specifications.cfm>
- [9] EVO videoconference: <http://evo.caltech.edu/evoGate/>

## 7.8 FEMTOSECOND PRECISION OPTICAL SYNCHRONIZATION

*F. Löhl, Cornell*

The ultra-short electron bunches used in soft and hard x-ray free-electron lasers (FEL) (see, e.g., [1]) significantly increase the demands on the precision with which timings in the accelerator have to be measured and controlled. Reference signals have to be distributed with femtosecond stability to various locations within the accelerator. There, these signals are used to synchronize radio-frequency (RF) systems (for example the RF fields in accelerating cavities), conventional lasers (e.g. for two-color pump-probe experiments in an FEL), or beam diagnostics, such as arrival-time monitors for the electron or photon beam.

Compared to conventional synchronization schemes in which RF signals are distributed via coaxial cables, the application of optical signals for transporting the synchronization signal has several advantages. Amongst these are the low-loss distribution in optical fibers, as well as the fact that optical methods potentially offer better time-resolution than RF-based methods, due to the many orders of magnitude higher frequency of optical signals, as well as due to the availability of ultra-short optical pulses. Two different optical synchronization schemes exist and will be described in the following. Both use lasers operating at a wavelength of around  $1.55 \mu\text{m}$ .

**Pulsed optical synchronization** In this scheme (see Fig.1), a mode-locked, erbium-doped laser serves as the timing reference and produces optical pulses with a transform limited duration of around 100 to 200 fs. The timing information is encoded in the highly accurate repetition rate of the reference laser. To ensure the long-term stability of the laser repetition rate, the laser is phase-locked to an RF source with a bandwidth of several kilohertz. The integrated timing jitter at frequencies larger than the locking bandwidth is only a few femtoseconds or less for many mode-locked lasers (see, for example, [2]).

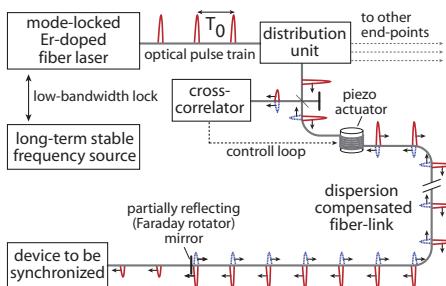


Figure 1: Schematic principle of the pulsed optical synchronization scheme (see text).

The laser pulses are distributed to remotely located end-stations via dispersion compensated optical fibers, which yields short pulses also at the end of the fiber-link. There, part of the laser power is reflected and travels back through the same fiber. Back at the distribution unit, the returning pulses are superimposed with laser pulses, which did not travel through the optical fiber, and are sent into a cross-correlator which measures changes of the temporal overlap between both pulse trains. By using a balanced optical cross-correlator as in refs.[3, 2], a dependency of the measured temporal overlap on laser power variations is suppressed and sub-femtosecond resolution is achieved.

A feedback loop based on the cross-correlator measurement compensates for variations of the pulse travel time through the optical fiber. Such variations can be induced, for example, by thermal expansion or contraction and by vibrations of the fiber, but also by less obvious causes like humidity variations which alter the mechanical properties of the fiber coating material. The actuator in this regulation loop is a piezo-electric fiber

stretcher or piezo-mirror in combination with an optical delay stage for larger timing drifts (not shown in Fig.1).

With an active feedback loop, the length of the optical fiber  $l_0$  is given as an integer multiple,  $m$ , of the laser repetition rate:  $l_0 = \frac{mc}{nf_0}$ . Here,  $n$  is the refractive index and  $f_0$  the laser repetition frequency. As a consequence, the relative length change of the optical fiber for a given laser frequency variation  $\Delta f$  is given by

$$\frac{\Delta l}{l_0} = \frac{l(f_0 + \Delta f) - l(f_0)}{l(f_0)} = \frac{f_0}{f_0 + \Delta f} - 1 \approx -\frac{\Delta f}{f_0} \quad \text{for } \Delta f \ll f_0 , \quad (1)$$

which determines the stability requirements of the frequency source to which the laser is phase-locked.

The laser pulses delivered via the optical fiber can be used for various applications the most important of which are:

**RF signal generation** The RF spectrum of a photo-diode signal, which detects the optical pulse train, contains harmonics at frequencies  $nf_0$ , with  $n = 1, 2, \dots$ , and the easiest way of extracting an RF signal from the optical pulse train is to select a desired harmonic with a bandpass filter. By this, an RF signal with sub-10 fs stability can be generated. However, for many photo detectors the timing of the extracted signal shifts when the optical power entering the detector varies. These drifts can be overcome by using an alternative scheme, which phase-locks the signal of a voltage controlled RF oscillator to the optical pulse train by performing an optical phase detection between both signals using a Sagnac loop interferometer [4].

**Laser synchronization** Mode-locked lasers can be synchronized to the optical pulse train by similar cross-correlation methods as they are used for the length stabilization of the optical fiber, thus also delivering sub-femtosecond precision. Examples of synchronizing two lasers of different center wavelengths with such a precision are given in refs.[4, 5].

**Electron bunch arrival-time measurement** The arrival times of single electron bunches can be measured with sub-10 fs resolution using an electro-optic detection scheme in which a beam pick-up signal is used to modulate the energies of the laser pulses inside of an electro-optical modulator [6]. The signal of this monitor can be used in a fast longitudinal beam feedback system to

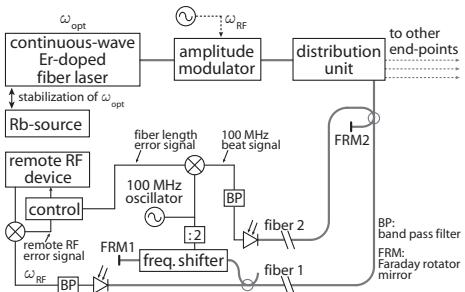


Figure 2: Schematic principle of the continuous wave optical synchronization scheme (see text).

stabilize the electron bunch arrival-time in superconducting accelerators [6].

**Continuous-wave optical synchronization** In a second optical synchronization scheme [7, 8], the signal of a narrow-band continuous-wave (CW) laser is distributed via optical fibers (see Fig.2). The laser operates at a wavelength of 1560.49 nm, which is stabilized by locking the laser frequency to a rubidium transition line. This is necessary, since the observed length of an optical fiber varies in an identical way with the laser frequency as stated in Eq.1. The laser signal is sent through an acousto-optical modulator in which the laser amplitude is modulated by the reference RF signal  $\omega_{RF}$ , which is to be distributed. The laser signal propagates through “fiber 1” to the remote location where the RF signal is extracted from the transmitted laser signal by photo-detection and subsequent band-pass filtering. Part of the laser power is split at the remote location and sent through an optical frequency shifter, in which the optical frequency ( $\sim 200$  THz) is up-shifted by  $\Delta f_s = 100$  MHz, before the laser beam is sent back through the same fiber. Back at the distribution unit, the returning signal is combined with a laser signal which did not travel through the optical fiber (reflected by the Faraday rotating mirror “FRM2”). The two superimposed signals propagate through a second fiber (“fiber 2”) back to the remote location, where the 100 MHz heterodyne beat signal between both laser beams is detected with a photo-diode and subsequent bandpass filtering. The phase of this beat signal changes by  $\pi$  when the phase of the returning, frequency shifted optical signal changes by  $\pi$ , which corresponds to a travel time difference of around 2.5 fs. A phase detection between the beat signal and the original 100 MHz RF signal

therefore yields a very sensitive measurement of “fiber 1” length variations.

It is possible to use piezo-electric actuators for the correction of fiber length variations similarly as described for the pulsed synchronization scheme. An alternative (shown in Fig.2) is to use a control loop which internally keeps track of fiber length variations without correcting them. As a second input, the control loop uses the signal of a phase detector measuring the timing between a remote RF device which needs to be synchronized, and the RF signal which was extracted from the distributed laser beam. Based on these two input signals, the control loop can adjust the phase of the remote RF device to keep it synchronized to the reference RF signal.

A difficulty in this scheme is that the phase velocity of the optical carrier differs from the group velocity of the RF signal, which is amplitude-modulated onto the optical carrier, due to chromatic dispersion. This leads to an additional correction that the control loop has to include. When this is done, the CW optical synchronization scheme can also provide a stability of the distributed reference RF signal of sub-10 fs [8].

*Laser synchronization* Mode-locked lasers can be synchronized to the CW optical reference with sub-femtosecond precision by stabilizing the carrier envelope offset frequency of the mode locked laser and then phase locking a frequency comb-line to the CW optical reference (see, for example, ref.[9]).

## References

- [1] Y. Ding et al, PRL 102, 254801 (2009)
- [2] J. Kim et al, Opt. Lett. 32, p. 3519 (2007)
- [3] J. Kim et al, Opt. Lett. 32, p. 1044 (2007)
- [4] J. Kim et al, Nature Photonics 2, p. 733 (2008)
- [5] T. R. Schibli et al, Opt. Lett. 28, p. 947 (2003)
- [6] F. Löhl et al, PRL 104, 144801 (2010)
- [7] Y. Sato et al, ALMA Memo No. 511 (2004)
- [8] R. Wilcox et al, Opt. Lett. 34, p. 3050 (2009)
- [9] A. Bartels et al, Opt. Lett. 28, p. 663 (2003)

# Chapter 8. RADIATION EFFECTS AND PROTECTION

## 8.1 RADIATION PROTECTION PRINCIPLES

*S. Roesler, M. Silari, CERN*

### 8.1.1 Quantities [1]

It would be desirable that the legal protection limits are expressed in directly measurable *physical quantities*. However, this does not allow to quantify biological effects of the exposure of the human body to ionizing radiation.

For this reason, protection limits are expressed in terms of so-called *protection quantities* which, although calculable, are not measurable. Protection quantities quantify the extent of exposure of the human body to ionizing radiation from both whole and partial body external irradiation and from intakes of radionuclides.

In order to demonstrate compliance with dose limits, so-called *operational quantities* are typically used which aim at providing conservative estimates of protection quantities. Often radiation protection detectors used for individual and area monitoring are calibrated in terms of operational quantities.

#### 8.1.1.1 Physical quantities

**Fluence**,  $\Phi$  (unit:  $1/m^2$ ) is the quotient of  $dN$  by  $da$ , where  $dN$  is the number of particles incident upon a small sphere of cross-sectional area  $da$ .

$$\Phi = dN/da . \quad (1)$$

In dosimetric calculations, fluence is frequently expressed in terms of the lengths of the particle trajectories. It can be shown that the fluence is also given by

$$\Phi = dl/dV , \quad (2)$$

where  $dl$  is the sum of the particle trajectory lengths in the volume  $dV$ .

**Absorbed dose**,  $D$  (unit: gray,  $1 \text{ Gy} = 1 \text{ J/kg} = 100 \text{ rad}$ ) is the energy imparted by ionizing radiation in a volume element of a specified material divided by the mass of this volume element.

**Kerma**,  $K$  (unit: Gy) is the sum of the initial kinetic energies of all charged particles liberated by indirectly ionizing radiation in a volume element of the specified material divided by the mass of this volume element.

**Linear energy transfer**,  $L$  or *LET* (unit:  $\text{J}/\text{m}$ , often given in  $\text{keV}/\mu\text{m}$ ) is the mean energy,  $dE$ , lost by a charged particle owing to collisions with electrons in traversing a distance  $dl$  in matter. *Low-LET radiation*: X-rays and gamma rays (accompanied by charged particles due to interactions with the surrounding medium) or light charged particles such as electrons that produce sparse ionizing events far apart at a molecular scale ( $L < 10 \text{ keV}/\mu\text{m}$ ). *High-LET radiation*: neutrons and heavy charged particles that produce ionizing events densely spaced at a molecular scale ( $L > 10 \text{ keV}/\mu\text{m}$ ).

**Activity**,  $A$  (unit: becquerel,  $1 \text{ Bq} = 1/\text{s} = 27 \text{ picocurie}$ ) is the expectation value of the number of nuclear decays occurring in a given quantity of material per unit time.

#### 8.1.1.2 Protection quantities

**Organ absorbed dose**,  $D_T$  (unit: Gy) in an organ or tissue  $T$  of mass  $m_T$  is defined as

$$D_T = \frac{1}{m_T} \int_{m_T} D dm . \quad (3)$$

**Equivalent dose**,  $H_T$  (unit: sievert,  $1 \text{ Sv} = 100 \text{ rem}$ ) in an organ or tissue  $T$  is equal to the sum of the absorbed doses  $D_{T,R}$  in the organ or tissue caused by different radiation types  $R$  weighted with so-called radiation weighting factors  $w_R$ :

$$H_T = \sum_R w_R \times D_{T,R} . \quad (4)$$

It expresses long-term risks (primarily cancer and leukemia) from low-level chronic exposure. The values for  $w_R$  recommended by the International Commission on Radiological Protection (ICRP) [2] are unity for photons, electrons and muons, 2.0 for protons and charged pions, 20.0 for ions and a function of energy for neutrons

$$\begin{aligned} E_n < 1 \text{ MeV} : \\ & 2.5 + 18.2 \times \exp[-(\ln E_n)^2/6] \\ 1 \text{ MeV} \leq E_n \leq 50 \text{ MeV} : \\ & 5.0 + 17.0 \times \exp[-(\ln(2E_n))^2/6] \\ E_n > 50 \text{ MeV} : \\ & 2.5 + 3.25 \times \exp[-(\ln(0.04E_n))^2/6]. \end{aligned} \quad (5)$$

**Effective dose**,  $E$  (unit: Sv) is the sum of the equivalent doses, weighted by the tissue weighting factors  $w_T$  ( $\sum_T w_T = 1$ ), of several organs and tissues  $T$  of the body that are considered to be most sensitive [2]

$$E = \sum_T w_T \times H_T. \quad (6)$$

### 8.1.1.3 Operational quantities

**Ambient dose equivalent**,  $H^*(10)$  (unit: Sv) is the dose equivalent at a point in a radiation field that would be produced by the corresponding expanded and aligned field in a 30 cm diameter sphere of unit density tissue at a depth of 10 mm on the radius vector opposing the direction of the aligned field. Ambient dose equivalent is the operational quantity for *area monitoring*.

**Personal dose equivalent**,  $H_p(d)$  (unit: Sv) is the dose equivalent in standard tissue at an appropriate depth,  $d$ , below a specified point on the human body. The specified point is normally taken to be where the individual dosimeter is worn. For the assessment of effective dose,  $H_p(10)$  with a depth  $d = 10$  mm is chosen, and for the assessment of the dose to the skin and to the hands and feet the personal dose equivalent,  $H_p(0.07)$ , with a depth  $d = 0.07$  mm, is used. Personal dose equivalent is the operational quantity for *individual monitoring*.

### 8.1.1.4 Dose conversion coefficients

Dose conversion coefficients allow direct calculation of protection or operational quantities from particle fluence and are functions of particle type, energy and irradiation configuration. The most common coefficients are those for effective dose

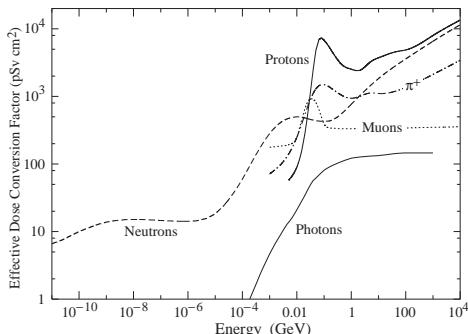


Figure 1: Fluence to effective dose conversion coefficients for antero-posterior irradiation and various particles [3].

and ambient dose equivalent. The former are based on simulations in which the dose to organs of anthropomorphic phantoms is calculated for approximate actual conditions of exposure, such as irradiation of the front of the body (antero-posterior irradiation) or isotropic irradiation.

Figure 1 shows, as an example, conversion coefficients from fluence to effective dose for antero-posterior irradiation. In Monte Carlo simulations such coefficients allow multiplication with fluence at scoring time such that effective dose to a human body at the considered location is directly obtained.

## References

- [1] Particle Data Group, "Review of Particle Physics" (2010), <http://pdg.lbl.gov>
- [2] ICRP Pub. 103, Annals of the ICRP, Elsevier (2007)
- [3] M. Pelliccioni, Radiat. Prot. Dosim. 88 (2000) 279

### 8.1.2 Health Effects of Ionizing Radiation

Radiation can cause two types of health effects, deterministic and stochastic:

**Deterministic effects** are tissue reactions which cause injury to a population of cells if a given threshold of absorbed dose is exceeded. The severity of the reaction increases with dose. The quantity in use for tissue reactions is the absorbed dose,  $D$ . When particles other than photons and electrons (low-LET radiation) are involved, a Relative Biological Effectiveness

(RBE)-weighted dose may be used. The RBE of a given radiation is the reciprocal of the ratio of the absorbed dose of that radiation to the absorbed dose of a reference radiation (usually X-rays) required to produce the same degree of biological effect. It is a complex quantity that depends on many factors such as cell type, dose rate, fractionation, etc.

**Stochastic effects** are malignant diseases and heritable effects for which the probability of an effect occurring, but not its severity, is a function of dose without threshold.

### 8.1.3 Radiation Levels

**Natural background radiation:** On a worldwide average, the annual whole-body dose equivalent due to all sources of natural background radiation ranges from 1.0 to 13 mSv with an annual average of 2.4 mSv [1]. In certain areas values up to 50 mSv have been measured. A large fraction (typically more than 50%) originate from inhaled natural radioactivity, mostly radon and radon decay products. Dose equivalent due to radon can vary by more than one order of magnitude: it is 0.1–0.2 mSv per year in open areas, 2 mSv per year on average in a house and more than 20 mSv per year in poorly ventilated mines.

**Cosmic ray background radiation:** At sea level the whole-body dose equivalent due to cosmic ray background radiation is dominated by muons, at higher altitudes also nucleons contribute. Dose equivalent rates range from less than 0.1  $\mu\text{Sv}/\text{h}$  at sea level to a few  $\mu\text{Sv}/\text{h}$  at aircraft altitudes.

**Cancer induction:** The cancer induction probability is about 5% per Sv on average for the entire population [2].

**Lethal dose:** The whole-body dose from penetrating ionizing radiation resulting in 50% mortality in 30 days, assuming no medical treatment, is 2.5–4.5 Gy (RBE-weighted when necessary), as measured internally on the body longitudinal center line. The surface dose varies due to variable body attenuation and may be a strong function of energy.

**Recommended dose limits:** The International Commission on Radiological Protection (ICRP) recommends a limit for radiation workers of 20 mSv effective dose per year averaged over 5 years, with the provision that the dose should

not exceed 50 mSv in any single year [2]. The limit in the EU-countries and Switzerland is 20 mSv per year, in the U.S. it is 50 mSv per year (5 rem per year). Many physics laboratories in the U.S. and elsewhere set lower limits. The dose limit for general public is typically 1 mSv per year.

## References

- [1] United Nations, Report of the United Nations Scientific Committee on the Effect of Atomic Radiation, General Assembly, Official Records A/63/46 (2008)
- [2] ICRP Pub. 103, Annals of the ICRP, Elsevier (2007)

## 8.2 SOURCES OF RADIATION

*R.H. Thomas, LLNL,  
updated by S. Roesler, M. Silari, CERN*

Radiations of most concern for radiological protection at particle accelerators are: (i) X-rays (bremsstrahlung) from electron or photon interactions; (ii) gamma-rays from neutron interactions; (iii) neutrons from either electron, photon, or nuclear interactions (iv) beta- and gamma-rays produced in the decay of radioactive materials; (v) muons at the highest energies.

Near the primary beam the production of photons dominates the radiological problems at electron accelerators. At proton accelerators hadrons primarily control the radiation field. In most practical radiation protection situations it is the interaction of the primary beam (and its secondary radiation, mainly neutrons) with thick objects (*e.g.*, beam-pipe walls, vacuum-system flanges, or Faraday cups) that is of concern. Such situations are usually described as “thick-target conditions”.

At a given distance  $r$  and at a specified angle  $\theta$ , information on secondary particle production (per primary particle) is given as differential fluence per unit solid angle (in units of  $\text{cm}^{-2} \text{sr}^{-1}$ )

$$\phi(\theta) = \frac{1}{r^2} \frac{dY(\theta)}{d\Omega} \quad (1)$$

or double differential fluence per unit solid angle and energy interval (in units of  $\text{cm}^{-2} \text{sr}^{-1} \text{MeV}^{-1}$ )

$$\frac{d\phi(E, \theta)}{dE} = \frac{1}{r^2} \frac{dY(E, \theta)}{dEd\Omega} \quad (2)$$

## Sec.8.2: SOURCES OF RADIATION

in which  $Y(\theta)$  is the integral yield per primary particle and  $dY(\theta)/d\Omega$  is the angular distribution of the yield.

The available angular distribution data for secondary-particle production in thin and thick targets have been compiled by the National Council on Radiation Protection and Measurements (NCRP) [1] but in some cases the data are quite uneven and must be used with appropriate caution. This section will mainly discuss the production of photons and neutrons.

### References

- [1] NCRP Report No. 144 (2003),  
<http://www.ncrppublications.org/>

#### 8.2.1 Electron Beams

For electron accelerators of all energies, bremsstrahlung dominates the secondary radiation field close to the primary beam loss point. Note, however, that at sufficiently high primary beam energies, shield thicknesses are usually determined by the production of neutrons of energy  $>20$  MeV. At energies  $>100$  MeV electromagnetic cascade processes become increasingly important in determining the external radiation field.

##### 8.2.1.1 Bremsstrahlung

Fig.1 [1] shows the bremsstrahlung efficiency for electrons stopped in various materials as a function of incident energy,  $E_0$ . The development of external bremsstrahlung as a function of target thickness is described by a “transition curve”: as the target thickness increases the radiation intensity initially increases until absorption in the target produces first a broad maximum and eventually declines with thickness in a roughly exponential manner. A target of thickness that yields the maximum radiation is called an “optimum” target and the photon spectrum emanating from such a target is described as a “thick-target bremsstrahlung spectrum”.

The dose rates produced by the more penetrating radiation from optimum, high  $Z$  targets bombarded by electrons of energy,  $E_0$  [MeV], may be estimated at zero degree (forward direc-

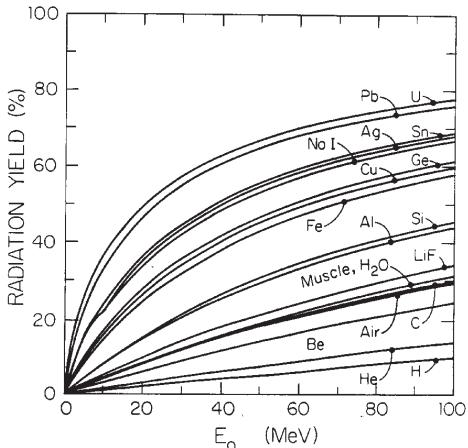


Figure 1: Bremsstrahlung efficiency for electrons stopped in various materials. The remaining energy is transferred to the medium by ionization and manifests itself as heat [1].

tion) as

$$\frac{dD}{dt} \approx 20 E_0^2 \quad (E_0 \leq 20 \text{ MeV}) \quad (1)$$

$$\frac{dD}{dt} \approx 300 E_0 \quad (E_0 > 20 \text{ MeV}) \quad (2)$$

where  $E_0$  is in MeV and the dose rate is in units of Gy/h at one metre from the target per kW of beam power on the target. At 90° (transverse direction) the dose rate is (in the same units)

$$\frac{dD}{dt} \approx 50 \quad (E_0 > 100 \text{ MeV}). \quad (3)$$

These yield equations are sufficiently accurate for shielding calculations.

The intensity of the thick-target bremsstrahlung in the forward direction is a slowly varying function with atomic number in the medium to high  $Z$  region. Furthermore, the bremsstrahlung intensity is increasingly forward-peaked with incident electron energy. In the energy range 2-20 MeV the measured data are adequately described by

$$E_0 \theta = 400 \quad [\text{MeV} \cdot \text{degree}] \quad (4)$$

where  $\theta$  is the angle relative to the incident beam for which the bremsstrahlung intensity is measured. For example, for  $E_0 = 10$  MeV about 15% of the radiation falls within 40 degrees and about 6% falls within 90° (see Fig.3.6 of Ref.[2]).

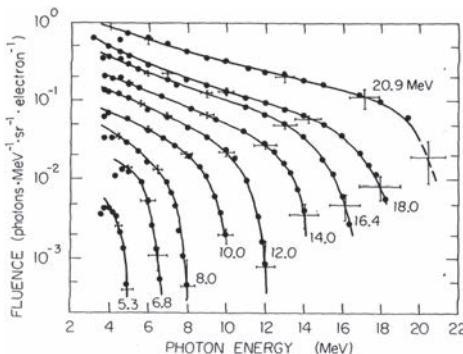


Figure 2: Bremsstrahlung spectra ( $0^\circ$ ) from  $0.2X_0$  targets of high  $Z$ .

At higher energies the behaviour is qualitatively similar and the angular width of the forward lobe of the distribution may be expressed in terms of the angle of half-intensity  $\theta_{1/2}$  by

$$E\theta_{1/2} = 100 \text{ [MeV} \cdot \text{degree]}. \quad (5)$$

For thin targets,  $X \ll X_0$ , the photon spectrum is

$$\frac{dN}{dE} \approx \frac{X}{X_0 E} \quad (6)$$

where  $E$  is the photon energy,  $X$  is the target thickness and  $X_0$  the radiation length.

The spectra from thick targets decline more rapidly than is the case for thin targets: at  $0^\circ$  this decline is roughly  $\propto E^{-2}$  and becomes steeper at larger angles. Examples of typical measured spectra are shown in Fig.2 [1]. The radiation spectra emerging from thick-target depend on target shape and material. Filtration by the target itself or by separate filters will alter the spectrum. The most satisfactory spectral calculations are by Monte-Carlo methods.

The photon field around high-energy electron accelerators results from the electromagnetic cascade developed in accelerator components and any intervening shielding. Outside substantial shielding two distinct radiation fields - the “broad field” and the “forward spike” - are observed.

The broad photon field is forward-peaked in the direction of the electron beam but extends with decreasing intensity to backward angles. This broad field is dominated by photons near the mass attenuation coefficient minimum of the absorbing material.

The very sharp forward-directed spike is a remnant of the radiation produced by the incident electrons and contains photons of the highest energy kinematically possible. The characteristic angle  $\theta_c$  of this radiation is  $\theta_c = m_e c^2 / E_0$ . For very thin targets,  $X \ll X_0$ , this spike will have the spectrum and other characteristics of thin-target bremsstrahlung. For thick targets, this spike may be observed above a background of photons from subsequent shower generations. The dose that might be delivered from the energy in this forward spike is considerable.

In the design of shielding for high-energy (multi-GeV) accelerators, the absorbed doses due to thick-target bremsstrahlung at large angles must be known. The dose outside the shield may be written as [2]

$$D(\theta) = 10^{-15} \left( 10.2 E_0 e^{-\theta^{0.6}} + 2.3 e^{-\theta/72} \right) \times E_0 \left( \frac{\sin \theta}{a+d} \right)^2 e^{-\frac{\mu}{\rho} \frac{\rho d}{\sin \theta}} \quad (7)$$

in which  $D$  is expressed in Gy per incident electron,  $E_0$  is the electron energy in GeV,  $\theta$  is the angle with respect to the beam direction,  $a$  is the internal radius of, e.g., a tunnel in meters,  $d$  is the transverse shield thickness around the tunnel in meters,  $\mu$  is the mass attenuation coefficient and  $\rho$  is the density of material.

Expressing Eq.(7) in terms of total incident electron energy  $U$  in Joules when  $D$  is in Gy, at  $\theta = 90^\circ$ , we have

$$D(90^\circ) = \frac{4.11 \times 10^{-6} U}{(a+d)^2} e^{-\frac{\mu}{\rho} \rho d}. \quad (8)$$

### 8.2.1.2 Electromagnetic cascade

For electrons in the GeV range and above, the electromagnetic cascade controls the distribution of energy in materials around the beam. Among others, an electromagnetic shower can be characterized by the quantities “radiation length” and “critical energy” that are both defined in Chapter 3.3.1.

The dosimetric properties of an electromagnetic cascade shower may be summarized by curves (Fig.3) showing the quantity of interest such as particle fluence or absorbed dose versus shower depth or distance from the shower axis. This curve may be generalized by the parameter (see Ref.[2] and references therein)

$$\lambda_l = 325 (\ln Z)^{-1.73} (\ln E_0) \quad (9)$$

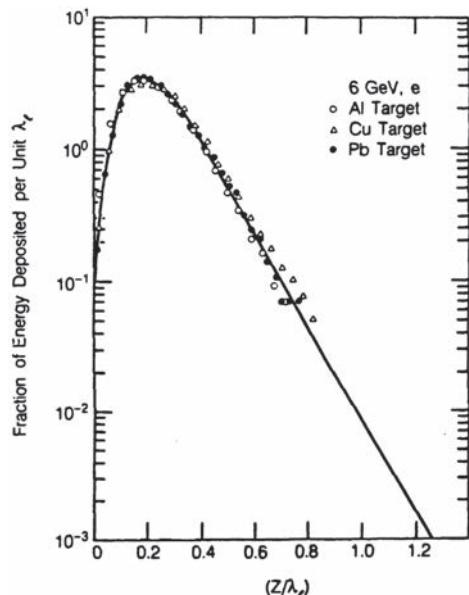


Figure 3: Fraction of total energy deposited by an electromagnetic cascade shower as a function of depth  $z$ , expressed in units of  $\lambda_l$ , integrated over all radii about the shower axis [2].

where  $\lambda_l$  is in  $\text{g/cm}^2$  and  $E_0$  is in MeV. When depths are expressed in units of  $\lambda_l$ , all curves approximately merge into a universal curve.

**Neutron production:** Photoneutron productions occurs via three mechanisms: the giant dipole resonance (from threshold up to about 35 MeV, the dominant process), the quasi-deuteron effect (30-300 MeV) and the neutrons released as a product of photo-pion reactions (above about 150 MeV). Neutron production takes place above a threshold value which varies from 10 to 19 MeV for light nuclei (with important exceptions, such as 2.23 MeV for deuterium and 1.67 MeV for beryllium) and from 4 to 6 MeV for heavy nuclei.

The giant dipole resonance reaction consists in a collective excitation of the nucleus, in which neutrons and protons oscillate in the direction of the photon electric field. The oscillation is damped by friction in a few cycles, with the photon energy being transferred to the nucleus in a process similar to evaporation. Nucleons emitted in the dipolar interaction have an anisotropic angular distribution, with a maximum at  $90^\circ$ , while those leaving the nucleus as a result of evapora-

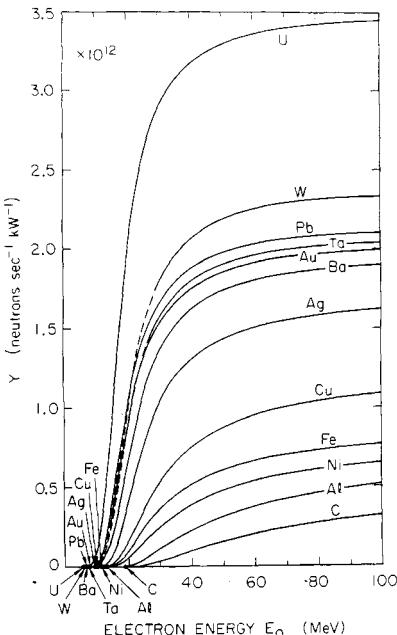


Figure 4: Photoneutron yield vs. incident energy for electrons on thick targets [1].

tion are emitted isotropically with a Maxwellian energy distribution described as [1]:

$$\frac{dN}{dE_n} = \frac{E_n}{T^2} \exp\left(-\frac{E_n}{T}\right) \quad (10)$$

where  $T$  is a nuclear 'temperature' (in units of MeV) characteristic of the particular target nucleus and its excitation energy. For heavy nuclei the 'temperature' generally lies in the range of  $T = 0.5 - 1.0$  MeV.

The dependence of photoneutron source strength  $N$  upon target material for electron beams of energy above 500 MeV is

$$N[\text{J}^{-1}] = 1.12 \times 10^8 Z^{0.66} \quad (11)$$

This formula is reliable over a wide range of  $Z$ , but underestimates the source for very light materials (especially  ${}^3\text{H}$ , Li and Be) and for transuranic materials in which photo-fission processes become important. For lower energies see Fig.4 [1].

## References

- [1] W.P. Swanson, IAEA Technical Report Series No. 188 (1979)
- [2] NCRP Report No. 144 (2003)

### 8.2.2 Proton Beams

The ionization range of protons increases with kinetic energy but their interaction mean free path becomes constant at high energies. Fig.1 shows both the ionization-range of protons in various materials and their probability of inelastic interaction within that range, as a function of proton energy. For radiation protection purposes “thick” target yields are needed: such yields have been measured directly but in many cases are determined from thin target cross section data by integrating over the range of the incident proton.

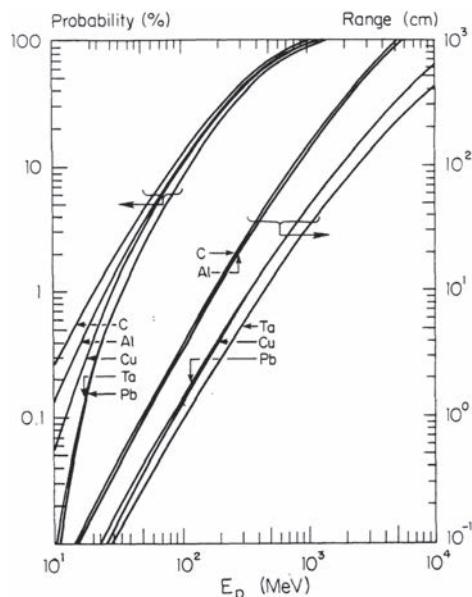


Figure 1: Proton range and probability of inelastic nuclear interaction within the range [1].

Neutrons typically present the largest radiation hazard at proton accelerators above  $\sim 10$  MeV. Modern accelerators produce high proton intensities and shielding must be provided to attenuate the neutron field. Photons and charged particles are preferentially removed by the shield, and the total dose rate outside of the shield is dominated by the neutrons. Thus, at proton accelerators at nearly all energies, it is important to understand the mechanisms of production and the yields of neutrons [2, 3, 4, 5].

Fig.2 of Ref.[3] summarizes the total yields per incident proton for several target materials

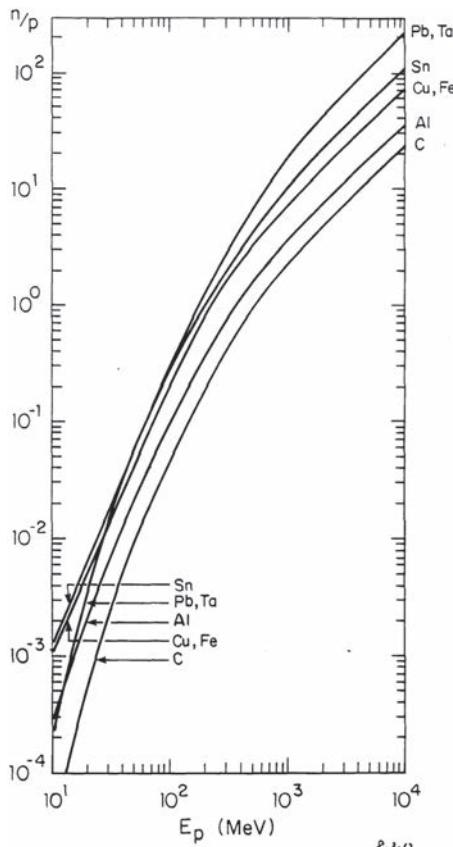


Figure 2: Total neutron yield per proton interacting in a thick target, for different target materials vs. energy [3].

as a function of incident proton energy up to a few GeV. The target thickness is comparable to the ionization range of protons (“thick target”). Between 50 and 500 MeV the neutron yield increases approximately as  $E_p^2$  while above 1 GeV it is approximately proportional to  $E_p$ . Details of target geometry may substantially influence the yield. This is particularly the case for the angular distribution of secondaries because of varying particle buildup or self-absorption within the target. Estimates of neutron yields from Fig.2 are accurate to about a factor of two.

One important parameter at high energies is the removal mean free path for protons that is approximately equal to that for neutrons. The removal mean free path of neutrons is related to

## Sec.8.2: SOURCES OF RADIATION

Table 1 Proton removal mean free path

| Material         | density<br>[g/cm <sup>3</sup> ] | $\lambda_{\text{removal}}$<br>[g/cm <sup>3</sup> ] | $\lambda_{\text{removal}}$<br>[cm] |
|------------------|---------------------------------|--|------------------------------------|
| H <sub>2</sub>   | 9.0·10 <sup>-5</sup>            | 43.3   | 4.8·10 <sup>5</sup>                |
| Be               | 1.85                            | 55.5   | 30.03                              |
| Al               | 2.7                             | 70.6   | 26.15                              |
| Fe               | 7.87                            | 82.8   | 10.52                              |
| Cu               | 8.96                            | 85.6   | 9.55                               |
| Pb               | 11.35                           | 116.2  | 10.24                              |
| U                | 18.95                           | 117.0  | 6.17                               |
| Air              | 1.29·10 <sup>-3</sup>           | 62.0   | 4.81·10 <sup>4</sup>               |
| H <sub>2</sub> O | 1.00                            | 60.1   | 60.1                               |
| Concrete         | 2.5                             | 67.4   | 26.96                              |
| SiO <sub>2</sub> | 2.64                            | 67.0   | 25.38                              |
| poly-ethylene    | 0.93                            | 56.9   | 61.51                              |

the neutron absorption cross section and becomes nearly independent of energy above a few hundred MeV. Table 1 summarizes removal mean free paths  $\lambda_{\text{removal}}$  (inelastic “attenuation length”), for protons in several common materials.

An empirical formula which fits measured angular distributions of the fluence of hadrons of energy  $> 40$  MeV at one metre from a copper target bombarded by protons in the energy region of  $5 < E_p < 500$  GeV is

$$\phi(\theta) = \frac{1}{2(\theta + 35\sqrt{E})^2} \quad (1)$$

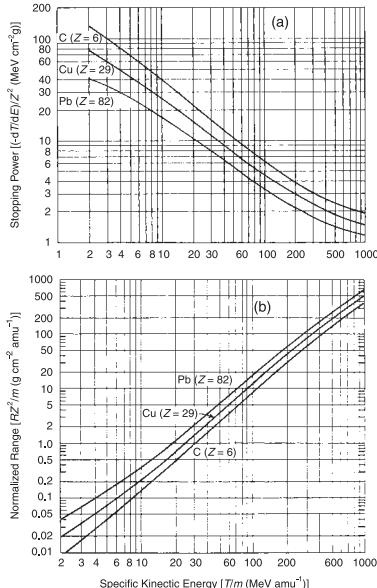
where  $\phi(\theta)$  is the fluence ( $\text{cm}^{-2}$ ) per interacting proton,  $E$  in GeV and  $\theta$  in degrees. Numerical integration of this formula yields a fairly good approximation to the total hadron flux emitted.

## References

- [1] NCRP Report No. 144 (2003)
- [2] A.H. Sullivan, A Guide to Radiation and Radioactivity levels Near High Energy Particle Accelerators, Nuclear Technology Pub. (1992)
- [3] R. Madey, Nuclear Accelerators in General, Section 2.2.2.1 in Engineering Compendium on Radiation Shielding, Vol.I, Springer (1968) p.49
- [4] H. W. Patterson, R.H. Thomas, Accelerator Health Physics, Academic (1973)
- [5] R.H. Thomas, G.R. Stevenson, IAEA Technical Report Series No. 283 (1988)

### 8.2.3 Ion Beams

There are some systematic differences in the variation of particle yields with mass of the projectile from reactions produced by light and heavier ions.



**Fig. 1.48.** Stopping power (a) and ranges (b) for protons in three materials. These curves can be used for other incident particles by adjusting for their atomic number ( $Z$ ) and mass [ $m$  (amu)]. The incident energy is thus expressed as the specific kinetic energy ( $T/m$ ). The curves are approximately correct except at the very lowest energies where charge exchange effects can be important and are probably most valid for  $m \leq 4$  (Engle, 1966; ICRU, 1993b).

©NCRP 2007 - All rights reserved.  
Licensed to CERN Library Interlibrary Loans  
Downloaded 06/29/07  
Single user license only, copying and networking prohibited.

Figure 1: Stopping power (a) and ranges (b) for protons in three materials. These curves can be used for other incident particles by adjusting for their atomic number ( $Z$ ) and mass [ $m$  (amu)]. The incident energy is thus expressed as the specific kinetic energy ( $T/m$ ) [2].

While the neutron yields from light-ion interactions show considerable fluctuations with projectile mass the yields produced by ions of medium to heavy mass are more nearly continuous functions of projectile mass. At very high energy (tens of GeV), neutron production from ions with mass number  $A$  is similar to that of  $A$  protons of the equivalent energy per nucleon.

The range of ions in matter can be calculated by the SRIM code [1]. Fig. 1 summarizes stopping power and range-energy relations for protons up to 1 GeV [2]. Data for light ions may be obtained

from these data by taking their atomic number  $Z$  and mass  $m$  (amu) into account. The curves are approximately correct except at the lowest energies where charge exchange effects can be important and are probably most valid for  $m \leq 4$ .

**Light ion reactions:** Deuteron stripping reaction ( $d,n$ ) usually have large neutron yields because the neutron of the deuteron is weakly bound (binding energy 2.225 MeV) and therefore easily stripped. The angular distribution is forward peaked. On average about half of the energy of the incident particle is invested in the outgoing neutrons with the energy spectrum centered upon roughly half of the incident kinetic energy. Many light ion reactions with targets of low mass number are exothermic with large cross sections. Such reactions can produce neutrons more energetic than the projectile.

**Heavy ion reactions:** There are now several intermediate-energy ( $< 1$  GeV/amu) heavy-ion accelerators operating worldwide. An extensive compilation of experimental data on double differential neutron production cross sections, thick target neutron production yields, neutron transport data and spallation products production cross-sections is provided by Ref.[3] for He to Xe ions in the energy range 100-800 MeV/nucleon.

For energies below 100 MeV/amu the total neutron yield  $Y$  (neutrons/ion) can be approximately fitted as a function of the energy per nucleon,  $E$  [MeV/amu],

$$Y(E, Z) = \frac{1.95 \times 10^{-4}}{Z^{2.75}} e^{-[0.475(\ln Z)^2]} E^{1.22\sqrt{Z}} \quad (1)$$

Eq.(1) shows that  $Y$  is not strongly dependent upon  $Z$ , the atomic number of the target. Yields are plotted in Fig.2 [4] where results are shown for incident ions spanning the domain from protons to very heavy ions up to the mass-region of lead. The shaded region gives an estimate of the probable error in yield estimates obtained by the use of Eq.(1).

**Neutron angular distribution** Ref.[5] gives a parametric representation of the neutron angular distribution valid over a wide range of energy per nucleon and for a number of different spectral conditions. The normalized angular distribution, of either dose equivalent or neutron fluence is

$$f(\theta, \xi) = \frac{1}{4\pi} \left[ \frac{1}{\log\left(1 + \frac{1}{\xi}\right)} \right] \left[ \frac{1}{\xi + \sin^2\left(\frac{\theta}{2}\right)} \right] \quad (2)$$

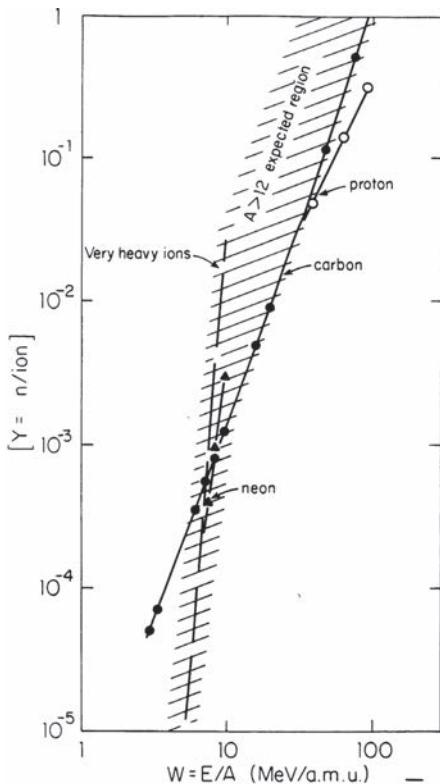


Figure 2: Total neutron yield vs specific energy. The shaded region is representative of the errors in the associated parametric fit given in the text [4].

with the fitting parameter

$$\xi \equiv \frac{\phi(90^\circ)}{\phi(0^\circ) - \phi(90^\circ)} \quad (3)$$

where  $\phi(\theta)$  is the value of the fluence or dose equivalent at  $\theta$ .

Values of the parameter  $\xi$  [4] are  $\xi = 0.07$  for uranium incident on uranium at 9 MeV/amu,  $\xi = 0.025$  for neutrons of energy below 20 MeV produced by 86 MeV/amu  $^{12}\text{C}$  incident on iron, and  $\xi = 3 \times 10^{-4}$  for neutrons of energy above 20 MeV produced by 86 MeV/amu  $^{12}\text{C}$  incident on iron. At higher energies the formulation of Eqs.(2) and (3) has not yet been tested.

At forward angles the maximum energy of the neutron spectrum extend up to about twice the energy per nucleon of the incident ion, whereas neutron spectra from protons extends up to the energy of the proton. Beyond a thick concrete shield,

### Sec.8.3: MONTE CARLO CODES

neutron spectra generated from high-energy ions (several GeV per nucleon) are qualitatively similar to neutron spectra from high-energy protons, with an evaporation component at around 1 MeV and a high-energy peak at around 100 MeV.

## References

- [1] J.F. Ziegler, J.P. Biersack, M.D. Ziegler, SRIM, The Stopping and Range of Ions in Matter, [www.srim.org](http://www.srim.org)
- [2] NCRP Report No. 144 (2003)
- [3] T. Nakamura, L. Heilbroon, Handbook on secondary particle production and transport by high-energy heavy ions, World Scientific (2006)
- [4] F. Clapier, C.S. Zaidins, Nucl. Instrum. Meth. 217 (1983) 489
- [5] V.E. Aleinikov et al., Radiat. Prot. Dosim. 11 (1985) 245

## 8.3 MONTE CARLO CODES

*S. Roesler, M. Silari, CERN*

The use of general-purpose particle interaction and transport Monte Carlo codes is often the most accurate and efficient choice for assessing radiation protection quantities at accelerators. Due to the vast spread of such codes to all areas of particle physics and the associated extensive benchmarking with experimental data, the modeling has reached an unprecedented accuracy. Furthermore, most codes allow the user to simulate all aspects of a high energy particle cascade in one and the same run: from the first interaction of a TeV nucleus over the transport and re-interactions (hadronic and electromagnetic) of the produced secondaries, to detailed nuclear fragmentation, the calculation of radioactive decays and even of the electromagnetic shower caused by the radiation from such decays.

In the following, short summaries on the most commonly used general-purpose Monte Carlo codes are given.

### 8.3.1 FLUKA

FLUKA is a general-purpose particle interaction and transport code with roots in radiation protection studies at high energy accelerators [1, 2]. It therefore comprises all features needed in this area of application, such as detailed hadronic and nuclear interaction models, full coupling between

hadronic and electromagnetic processes and numerous variance reduction options.

The module for hadronic interactions is called PEANUT and consists of a phenomenological description (Dual Parton Model-based Glauber Gribov cascade) of high energy interactions (up to 20 TeV), a generalized intranuclear cascade and pre-equilibrium emission models as well as models for evaporation, fragmentation, fission and de-excitation by gamma emission. Interactions of ions are simulated through interfaces with different codes based on models applicable in certain ranges of energy (DPMJET3 above 5 GeV/n, rQMD-2.4 between 0.1 and 5 GeV/n, Boltzmann Master Equation below 0.1 GeV/n) (see [2] and references therein).

The transport of neutrons with energies below 20 MeV is performed by a multi-group algorithm based on evaluated cross section data (ENDF/B, JEF, JENDL, etc.) binned into 260 energy groups, 31 of which in the thermal energy region. For a few isotopes ( $^1\text{H}$ ,  $^6\text{Li}$ ,  $^{10}\text{B}$ ,  $^{14}\text{N}$ ) pointwise cross sections can be optionally used during transport. The detailed implementation of electromagnetic processes in the energy range between 1 keV and 1 PeV is fully coupled with the models for hadronic interactions.

Many variance reduction techniques are available in FLUKA, including weight windows, region importance biasing, and leading particle, interaction, and decay length biasing (among others). The capabilities of FLUKA are unique for studies of induced radioactivity, especially with regard to nuclide production, decay, and transport of residual radiation. In particular, particle cascades by prompt and residual radiation are simulated in parallel based on the microscopic models for nuclide production and a solution of the Bateman equations for activity build-up and decay.

## References

- [1] A. Ferrari et al, CERN-2005-010 (2005), <http://www.fluka.org>
- [2] G. Battistoni et al, Proc. of the Hadronic Shower Simulation Workshop 2006, M. Albrow, R. Raja eds., AIP Conference Proceedings 896 (2007) 31

### 8.3.2 GEANT4

GEANT4 is an object-oriented toolkit originally designed to simulate detector responses of modern particle and nuclear physics experiments [1,

2]. It consists of a kernel that provides the framework for particle transport, including tracking, geometry description, material specifications, management of events and interfaces to external graphics systems.

The kernel also provides interfaces to physics processes. In this regard, the flexibility of GEANT4 is unique as it allows the user to freely select the physics models that best serve the particular application needs. Implementations of interaction models exist over an extended range of energies, from optical photons and thermal neutrons to high-energy interactions required for the simulation of accelerator and cosmic ray experiments. In many cases, complementary or alternative modeling approaches are offered from which the user can choose.

Descriptions of intra-nuclear cascades include implementations of the Binary and the Bertini cascade models. Both are valid for interactions of nucleons and charged mesons, the former for energies below 3 GeV, and the latter for energies below 10 GeV. At higher energies (up to 10 TeV), three models are available: a high-energy parameterized model (using fits to experimental data), a quark-gluon string model, and the Fritiof fragmentation model, all three based on string excitations and decay into hadrons. Nuclear de-excitation models include abrasion-ablation and Fermi-breakup models. Furthermore, heavy-ion interactions can also be simulated if the appropriate packages are linked.

The package for electromagnetic physics comprises the standard physics processes as well as extensions to energies below 1 keV, including emissions of X-rays, optical photon transport, *etc.*

To facilitate the use of variance reduction techniques, general-purpose biasing methods such as importance biasing, weight windows, and a weight cut-off method have been introduced directly into the toolkit. Other variance reduction methods, such as leading particle biasing for hadronic processes, come with the respective physics packages.

## References

- [1] S. Agostinelli et al, Nucl. Instrum. Meth. A506 (2003) 250
- [2] J. Allison et al, IEEE Transactions on Nuclear Science 53 (2006) 270

### 8.3.3 MARS15

The MARS15 code system [1, 2] is a set of Monte Carlo programs for the simulation of hadronic and electromagnetic cascades that is used for shielding, accelerator design, and detector studies. Correspondingly, it covers a wide energy range: 1 keV to 100 TeV for muons, charged hadrons, heavy ions and electromagnetic showers; and 0.00215 eV to 100 TeV for neutrons.

Hadronic interactions above 5 GeV can be simulated with either an inclusive or an exclusive event generator. While the former is CPU-efficient (especially at high energy) and based on a wealth of experimental data on inclusive interaction spectra, the latter provides final states on a single interaction level and preserves correlations. In the exclusive mode, the cascade-exciton model CEM03.03 describes hadron-nucleus and photon-nucleus interactions below 5 GeV, the Quark-Gluon String Model code LAQGSM03.03 simulates nuclear interactions of hadrons and photons up to 800 GeV and of heavy ions up to 800 GeV/nucleon, and the DPMJET3 code treats the interactions at higher energies. The exclusive mode also includes models for a detailed calculation of nuclide production via evaporation, fission, and fragmentation processes.

MARS15 is also coupled to the MCNP4C code that handles all interactions of neutrons with energies below 14 MeV. Produced secondaries other than neutrons are directed back to the MARS15 modules for further transport.

Different variance reduction techniques, such as inclusive particle production, weight windows, particle splitting, and Russian roulette, are available in MARS15. A tagging module allows one to tag the origin of a given signal for source term or sensitivity analyses. Further features of MARS15 include a MAD-MARS Beam-Line Builder for a convenient creation of accelerator models.

## References

- [1] N.V. Mokhov, S.I. Striganov, Proc. of the Hadronic Shower Simulation Workshop 2006, M. Albrow, R. Raja eds., AIP Conference Proceedings 896 (2007)
- [2] N.V. Mokhov, "MARS Code System", Version 15 (2009), <http://www-ap.fnal.gov/MARS>

### 8.3.4 MCNPX

MCNPX originates from the Monte Carlo N-Particle transport (MCNP) family of neutron interaction and transport codes and, therefore, features one of the most comprehensive and detailed descriptions of the related physical processes [1, 2]. Later it was extended to other particle types, including ions and electromagnetic particles. This allowed an expansion of the areas of application from those purely neutronics-related to accelerator shielding design, medical physics, and space radiation, among others.

The neutron interaction and transport modules use standard evaluated data libraries mixed with physics models where such libraries are not available. The transport is continuous in energy and includes all features necessary for reactor simulations, including burn-up, depletion, and transmutation. Different generalized intra-nuclear cascade codes can be linked to explore different physics implementations, such as CEM2K, INCL4 and ISABEL (see [2] and references therein). They either contain fission-evaporation models or can be coupled to such models (*e.g.*, ABLA), allowing detailed predictions for radionuclide production. While the intra-nuclear cascade codes are limited to interaction energies below a few GeV, a link to the Quark-Gluon String Model code LAQGSM03 extends this energy range to about 800 GeV. The latter code also allows the simulation of ion interactions. Electromagnetic interactions are simulated in MCNPX by the ITS 3.0 code.

MCNPX contains one of the most powerful implementations of variance reduction techniques. Spherical mesh weight windows can be created by a generator in order to focus the simulation time on certain spatial regions of interest. In addition, a more generalized phase space biasing is also possible through energy- and time-dependent weight windows. Other biasing options include pulse-height tallies with variance reduction and criticality source convergence acceleration.

### References

- [1] D.B. Pelowitz, ed., Los Alamos National Laboratory report, LA-CP-05-0369 (2005)
- [2] G. McKinney et al, Proc. of the International Workshop on Fast Neutron Detectors, University of Cape Town, South Africa (2006)

### 8.3.5 PHITS

The Particle and Heavy-Ion Transport code System PHITS (see [1, 2] and references therein) was among the first general-purpose codes to simulate the transport and interactions of heavy ions in a wide energy range, from 10 MeV/nucleon to 100 GeV/nucleon. It is based on the high-energy hadron transport code NMTC/JAM that was extended to heavy ions by incorporating the JAERI Quantum Molecular Dynamics code JQMD.

Below energies of a few GeV, hadron-nucleus interactions in PHITS are described through the production and decay of resonances, while at higher energies (up to 200 GeV) inelastic hadron-nucleus collisions proceed via the formation and decay of so-called strings that eventually hadronize through the creation of (di)quark-anti(di)quark pairs. Both are embedded into an intra-nuclear cascade calculation. Nucleus-nucleus interactions are simulated within a molecular dynamics framework based on effective interactions between nucleons.

The generalized evaporation model GEM treats the fragmentation and de-excitation of the spectator nuclei and includes 66 different projectiles (up to Mg) and fission processes. The production of radioactive nuclides, both from projectile and target nuclei, thus follows directly from the mentioned microscopic interaction models.

The transport of low-energy neutrons employs cross sections from evaluated nuclear data libraries such as ENDF and JENDL below 20 MeV and LA150 up to 150 MeV. Electromagnetic interactions are simulated based on the ITS code in the energy range between 1 keV and 1 GeV. Several variance reduction techniques, including weight windows and region importance biasing, are available in PHITS.

### References

- [1] H. Iwase, K. Niita and T. Nakamura, Journal of Nuclear Science and Technology 39 (2002) 1142
- [2] K. Niita et al, Radiation Measurements 41 (2006) 1080

## 8.4 RADIATION TRANSPORT AND SHIELDING

*S. Roesler, M. Silari, CERN*

### 8.4.1 General Concepts

The basic procedure to be followed for a shielding assessment consists of a 3-step approach: 1) the determination of the strength of the radiation source (or source term), which depends on various parameters such as type and energy of accelerated particle, beam current, distribution and intensity of beam losses, and so on, 2) the specification of the maximum admissible ambient dose equivalent levels outside the shielding and 3) the design of a shield with adequate attenuation to meet the ambient dose equivalent limits. For both neutrons and photons, the attenuation of radiation by a shield can usually be approximated by an exponential function. This section discusses shielding of neutrons, which is the dominant radiation for which barriers are to be designed at proton accelerator facilities. Neutrons also represent the more penetrating radiation at high-energy accelerators, where otherwise bremsstrahlung X-rays dominate. Muon shielding only becomes important at energies larger than about 10 GeV and determines the shielding requirements in the forward direction at very high energies, so that in fact the only practical solution for forward shielding at very high-energy accelerators is to place the machine underground and make use of soil as shielding material.

The present state-of-the-art Monte Carlo radiation transport codes (see Chapter 8.3) along with the extensive computing power available nowadays are the best tools to design shielding in very complex geometrical conditions. Nonetheless it is often useful – although not always possible – to perform a first assessment of the required shielding thickness by using a simple model, to be verified at a more advanced stage of the project with a Monte Carlo simulation in a more realistic geometry of the facility. Often this simplified approach provides a conservative estimate of the required shielding.

**Beam loss assumptions.** There are typically two situations which have to be considered when designing a shield. The first is routine accelerator operation, for which one can expect (more or less) constant losses at (more or less) well defined locations in the machine such as septa, collimators,

dumps and aperture restrictions. The second is that of an accidental scenario with full loss of the beam at a defined location or over a given section of the accelerator. In the former case one has to design the shield to meet the requirements of ambient dose equivalent rate according to the classification of the area past the shield. In the latter case one has to assess the dose rate  $dH/dt$  in the accessible area generated by the beam loss, define the maximum acceptable integrated dose  $H_{\max}$  that can be received by the personnel who may be present in the area (typically from tens of  $\mu\text{Sv}$  to a few mSv, according to the likelihood of the event), and then make sure that the accidental condition cannot last for a time  $T$  longer than  $T = H_{\max}/(dH/dt)$ . This in turn requires that the accelerator control system must be capable of stopping the beam within the time  $T$ . If the loss is “catastrophic”, that is if it cannot continue over a certain time, but it occurs in a single shot at a defined point and will cause such a damage to the accelerator that the beam is automatically stopped, then one has to evaluate the integral dose generated by such event as a function of shielding thickness, and dimension the barrier accordingly. The final shielding thickness is the largest value imposed by the two conditions, routine and accidental. It is often found that the former dictates the shielding design.

### 8.4.2 Semi-Empirical Models

In a number of practical cases and over a broad range of shield thicknesses, a simple point kernel equation is sufficiently accurate to estimate shielding (see, for example, Refs.[1, 2]):

$$H(E_p, \theta, d/\lambda) = \frac{H_0(E_p, \theta)}{r^2} \exp \left[ -\frac{d}{\lambda(\theta) \cos(\alpha)} \right] \quad (1)$$

where  $H$  is the ambient dose equivalent beyond the shield,  $E_p$  is the proton energy,  $r$  is the distance between the radiation source (the target stopping the protons) and the scoring position,  $\theta$  is the angle between the direction of the emitted radiation and the beam axis,  $H_0$  is the source term depending on the target material,  $d$  is the shield thickness,  $\lambda(\theta)$  is the attenuation length for the given shielding material at emission angle  $\theta$ , and  $\alpha$  is the angle between the direction of the emitted radiation and the normal to the shield surface. Below about 1 GeV,  $\lambda$  varies strongly with neutron energy and depends on the target material.

At energies above 1 GeV and for transverse shielding,  $\theta = 90^\circ$  (high-energy accelerators are usually installed in a tunnel rather than in a hall, so that in most cases one has to evaluate transverse shielding), the Moyer model [3] can be employed, for which  $H_0(E_p, 90^\circ) = 1.26 \times 10^{-14} E_p^{0.8}$  Sv m<sup>2</sup> (with  $E_p$  in GeV) and  $\lambda = 118$  g cm<sup>-2</sup> for concrete or earth and  $\lambda = 167$  g cm<sup>-2</sup> for iron.  $H_0(E_p, 90^\circ)$  is now a slowly varying function of incident proton energy and is essentially independent of target material. There also exists a formulation of the Moyer model for an infinite line-source [4, 5], valid for proton energies above about 1 GeV. For one proton per metre interacting uniformly over the whole length of an infinite source, the dose equivalent past the shield is given by:

$$\begin{aligned} H(r) &= \frac{\psi}{r} \int_0^\pi \exp(-\beta\theta) \exp\left(-\frac{d}{\lambda \sin\theta}\right) d\theta \\ &= \frac{\psi}{r} M(\beta, d/\lambda) \end{aligned} \quad (2)$$

The parameter  $\beta$  has the value  $\beta = 2.3$  for  $\theta = 90^\circ$ .  $\psi = 2.84 \times 10^{-13} E_p^{0.8}$  Sv m<sup>2</sup> is the source term. The integral in this equation,  $M(\beta, d/\lambda)$ , is known as the Moyer integral. Values of the Moyer integral  $M(2.3, d/\lambda)$  are given in Table 2.25 of Ref.[5].<sup>1</sup>

Point-source line-of-sight models can be directly applied to shielding proton accelerators of energy up to about 500 MeV, where the radiation source generated by the interacting protons is limited to a distance equal to the proton range. At higher energies the cascade does not generally develop in the target and many of the high-energy secondaries can escape and generate a cascade in the shielding. In addition the cascade extends over a comparatively long distance as compared to the dimension of an accelerator tunnel, and thus can no longer be regarded as a point source. In such cases the wall-source model described in Ref.[6] may be more appropriate.

## References

- [1] R.H. Thomas, G.R. Stevenson, IAEA Technical Report Series No. 283 (1988)
- [2] NCRP Report No. 144 (2003)
- [3] B.J. Moyer, Lawrence Radiation Laboratory Report UCRL-9769 (1961)
- [4] J.T. Routti et al., Nucl. Instrum. Meth. 76 (1969) 157
- [5] A. Fassò et al., Landolt-Börnstein, Volume 11, Ed. H. Schopper, Springer-Verlag, Berlin (1990).
- [6] G.R. Stevenson, Radiat. Prot. Dosim., 96 (2001) 359

### 8.4.3 The Monte Carlo “hybrid” Method

A hybrid approach between a point-source line-of-sight model and a full Monte Carlo treatment consists in the determination by Monte Carlo simulations of source terms and attenuation lengths as a function of target, material, beam energy and emission angle, to be applied in generic shielding situations with a simple point-kernel model [1, 2]. It must be stressed that the “source term” which appears in Eq.(1) of Chapter 8.4.2 and in the Moyer model does not refer to a bare source but to a virtual source derived from extrapolating the dose equivalent deep in the shield to zero thickness. Therefore, it must not be used to estimate the dose rate from an unshielded source or as a source term for evaluating transmission of radiation through a duct or a maze. It is also worth pointing out that the simplified Eq.(1) does not show the build-up term present at shallow depths in the shield [1, 3, 4].

In general, the attenuation of the total ambient dose equivalent through a thick shield can be expressed through the classical two-parameter formula given by Eq.(1) of Chapter 8.4.2. There are however a number of cases in which the attenuation of the dose equivalent is better expressed by a double-exponential function:

$$\begin{aligned} H(E_p, \theta, d/\lambda) &= \\ &\frac{H_1(E_p, \theta)}{r^2} \exp\left[-\frac{d}{\lambda_1(\theta) \cos(\alpha)}\right] + \\ &\frac{H_2(E_p, \theta)}{r^2} \exp\left[-\frac{d}{\lambda_2(\theta) \cos(\alpha)}\right] \end{aligned} \quad (1)$$

where  $H_1$ ,  $\lambda_1(\theta)$  and  $H_2$ ,  $\lambda_2(\theta)$  are the source terms and the attenuation lengths for the shallow-depth and large-depth (deep penetration) exponential functions, respectively (see, for example, Refs.[1, 2]). In practice, the above equation includes Eq.(1) of Chapter 8.4.2 by setting  $H_0 = H_2$ ,  $\lambda(\theta) = \lambda_2(\theta)$  and setting the first term to zero (*i.e.*,  $H_1 = \lambda_1(\theta) = 0$ ).

This expression can be equally used to estimate shielding of neutrons generated by proton and ion beams of intermediate energy striking thick-targets of various materials. Source terms

<sup>1</sup>The parameter  $\beta$  in Eq.2 is not to be confused with the relativistic velocity  $\beta = v/c$ .

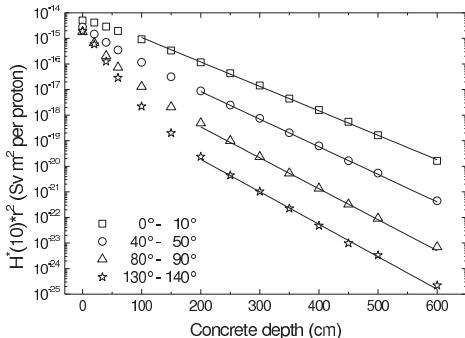


Figure 1: Attenuation of total dose equivalent in concrete for 250 MeV protons on a thick iron target at four intervals of the emission angle. The statistical uncertainties are smaller than the size of the symbols (from Ref.[1]).

and attenuation lengths (for shallow-depth and deep penetration), mostly for concrete and some for iron, can be found in the literature (see, for examples, Refs.[1, 2, 5] and references therein). As an example, the attenuation of the total dose equivalent in ordinary concrete at four emission angles is plotted in Fig.1 for 250 MeV protons impinging on a thick iron target.

At proton accelerators the total ambient dose equivalent past a concrete shield is mainly due to neutrons, but photons and protons sometimes cannot be completely neglected. The non-neutron part of the dose equivalent is usually limited to a few percent, except in the backward direction where photons contribute more than 10% and up to 35% to the total dose for a shield thickness of one to two meters. Generally speaking, the non-neutron contribution decreases with increasing energy of the primary proton and increases with increasing angle: in particular, the contribution from secondary protons is maximum in the forward direction (because of protons generated in the intranuclear cascade) and decreases with angle, whereas the electromagnetic contribution (possibly dominated by prompt gamma rays) moves in the opposite direction.

It should be underlined that the target thickness plays a fundamental role for shielding. Generally, for a thin target, shielding data are conservative, since the outgoing neutron spectrum is much harder and thus penetrating, the attenuation length is longer and no-self absorption is provided

by the target itself. In contrast, the secondary particle yield is higher for a thick target, showing a more intense low-energy component. Usually, in a practical case the neutron-producing target is rather thick (e.g., beam loss in a magnet or in a collimator) and thick-target data should normally be used.

## References

- [1] S. Agosteo et al., Nucl. Instrum. Methods B265 (2007) 581
- [2] S. Agosteo et al., Nucl. Instrum. Methods B266 (2008) 3406
- [3] R.H. Thomas, G.R. Stevenson, IAEA Technical Report Series No. 283 (1988)
- [4] G.R. Stevenson, Radiat. Prot. Dosim. 359 (2001) 359
- [5] S. Agosteo et al., Nucl. Instrum. Meth. B217 (2004) 221

### 8.4.4 Photon Transmission

Photons dominate the shielding design at low energy electron accelerators. Most of the electron accelerators operate in the energy range of 1 – 50 MeV and are used in cancer radiation therapy, industrial processes and radiography. To produce photons the electron beam is usually stopped in a high- $Z$  target. Photon emission becomes increasingly forward directed with increasing electron energy. The attenuation factor  $B$  of a primary barrier (a shield towards which the bremsstrahlung beam can be directed) is given by [1, 2]:

$$B = \dot{H}^*(10)d^2/(WUT) \quad (1)$$

in which  $\dot{H}^*(10)$  is the limiting value of the ambient dose equivalent per unit time (e.g., Sv/week),  $d$  is the distance between the source and the point where  $H^*(10)$  is evaluated (m),  $W$  is the so-called workload (e.g., Gy m<sup>2</sup>/week),  $U$  is the use factor of the barrier and  $T$  is the occupancy factor of the area past the shield.  $WUT$  represent the source term. Transmission curves in various materials are given in Refs.[1, 2].

Secondary barriers are designed to shield scattered radiation. For their design use is often made of the albedo coefficients similarly to the labyrinth design discussed in Chapter 8.4.6. A thorough discussion of photon shielding including design of primary and secondary barriers and shielding against giant-resonance neutrons is given in Refs.[1, 2].

## References

- [1] W.P. Swanson, IAEA Technical Report Series No. 188 (1979)
- [2] NCRP Report No. 144 (2003)

### 8.4.5 Ducts and Labyrinths for Neutron Attenuation

Ducts and labyrinths traversing the shielding serve as access path for control and power cables, ventilation pipes, waveguides for RF power, cooling and cryogenic lines, as well as for personnel and equipment access. At the same time they represent a leakage path for radiation, especially neutrons, and must be properly designed in order not to compromise the global effectiveness of the shield. There are two general rules which should be obeyed in the design of adequate radiation attenuation of penetrations:

- Shield-penetrations should never be placed so that the primary beam can point directly towards it or in any location that permits an unshielded leakage path for the primary beam or produced secondary radiation.
- For any adequate labyrinth, the sum of the shield-wall thicknesses between the source of radiation and the exit point of the penetration should be at least equivalent to that which would be required if the labyrinth were not present.

Monte Carlo simulations are the best tool for calculating radiation streaming through ducts and labyrinths. However, the design of labyrinths with right-angle bends can often be achieved to a good level of accuracy by simple analytical expressions or universal transmission curves, so-called as the depth in a duct or in a maze leg  $d$ , is normalized to its cross-sectional area  $A$ , and thus expressed as a function of  $d/\sqrt{A}$ .

The universal transmission curves of Figs.1 and 2 [1, 2] provide the transmission of ambient dose equivalent for a straight duct or a first leg of a labyrinth, and for the second and subsequent legs of a labyrinth, respectively.

A parametric form of the transmission curves has been given by Stevenson and Fassò [3] for the case of an off-axis source. The transmission  $T$  in the first and second leg can be expressed as:

$$T = 1/(1 + 2.5\sqrt{D} + 0.17D^{1.7} + 0.79D^3) \quad (1)$$

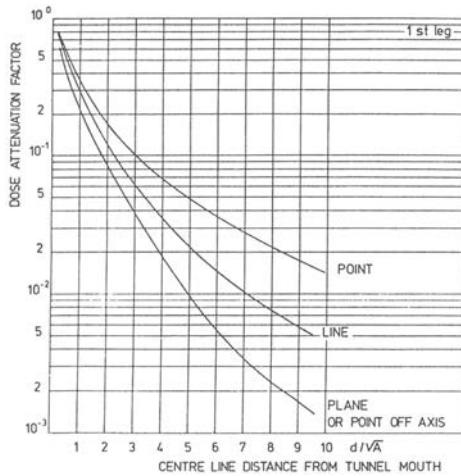


Figure 1: Universal transmission curves for the first leg of a labyrinth (from Ref.[1]).

and

$$T = 1/[1 + 2.8D(1.57)^{d+2}] \quad (2)$$

where  $D = d/\sqrt{A}$ .

The geometric and angular distribution of the neutrons source and its position with respect to the maze mouth, considerably affect the attenuation factor provided by the first leg of the maze. However, after the first bend the neutrons essentially lose “memory” of their original spectrum. For the second and subsequent legs of a labyrinth the position of the source (plane or point off-axis, linear or point on-axis) is no longer relevant. This reflects into a single universal curve (Fig.2). The energy of the proton beam causing the neutron emission is also not much relevant when considering the attenuation provided by the second and subsequent legs of the maze, especially for emission at 90 degrees.

Whereas the curves in Fig.1 for a line and plane or point off-axis source give correct predictions of the dose attenuation, the accuracy of the curve for a point source on-axis, *i.e.*, a localized beam loss in direct view of the duct aperture, is much lower. The neutron transmission through a straight duct of length  $d$  and cross-sectional area  $A$  in direct view of the source can instead be estimated by the expression [4]:

$$T = (k\sqrt{A})(d/\sqrt{A})^{-2h\sqrt{A}} \quad (3)$$

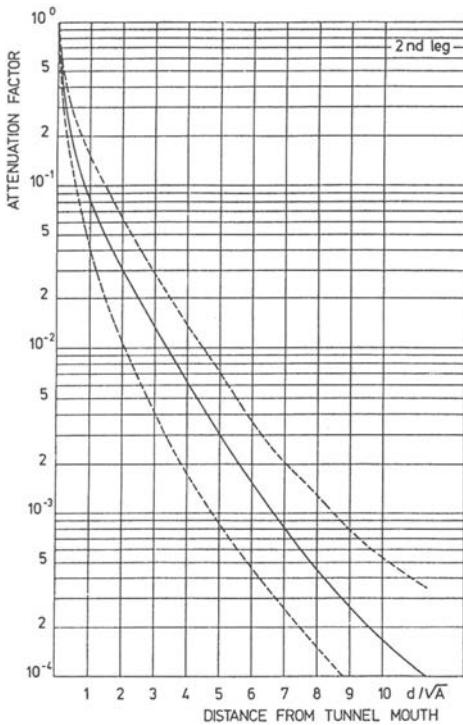


Figure 2: Universal transmission curves for the second and subsequent legs of a labyrinth (from Ref.[1]). The dashed lines indicate appropriate confidence limits, as explained in Refs.[1, 2].

Table 1: Fit parameters  $k_1, k_2, h_1, h_2$  to be used in Eq.(4).

| $k_1$  | $k_2$  | $h_1$  | $h_2$  |
|--------|--------|--------|--------|
| 1.7787 | 1.2728 | 0.8304 | 0.7307 |

with the parameters  $k$  and  $h$  given by

$$k = k_1 A^{-k_2/2}, \quad h = h_1 A^{-h_2/2} \quad (4)$$

with  $k_1, k_2, h_1, h_2$  given in Table 1, for any given cross-sectional area  $A$  of the duct. The exponential in Eq.(3) quantifies the deviation from the inverse-square law. For typical straight ducts with diameters as given in Table 2, Eq.(3) can be used directly with the parameters of Table 3. The same expression can be used for non-circular ducts with equivalent cross-section area.

Table 2: Typical diameters for circular ducts and mazes according to their use.

| Use                              | Diam. (cm) |
|----------------------------------|------------|
| Cables, cooling, cryogenics      | 30         |
| RF waveguides, cooling, cryo.    | 60         |
| RF waveguides, ventilation       | 110        |
| Access of personnel and services | 225        |
| Access of personnel and equipm.  | 340        |

Table 3: Fit parameters  $k$  and  $h$  calculated for the ducts of Table 2.

| $d$ (cm) | $k$ ( $\text{m}^{-1}$ ) | $h$ ( $\text{m}^{-1}$ ) |
|----------|-------------------------|-------------------------|
| 30       | 4.18                    | 2.38                    |
| 60       | 1.70                    | 1.18                    |
| 110      | 0.76                    | 0.78                    |
| 225      | 0.39                    | 0.56                    |
| 340      | 0.17                    | 0.36                    |

## References

- [1] K. Goebel et al., CERN Lab II-RA/Note/75-10 (1975)
- [2] R.H. Thomas, G.R. Stevenson, IAEA Technical Report Series No. 283 (1988)
- [3] G.R. Stevenson, A. Fassò, CERN Divisional Report TIS-RP/185/CF (1987)
- [4] E. Mauro, M. Silari, Nucl. Instrum. Meth. A608 (2009) 28
- [5] NCRP Report No. 144 (2003)

### 8.4.6 Ducts and Labyrinths for Photon Attenuation

Neutron attenuation data *cannot* be used to estimate the attenuation of electromagnetic radiation. The same duct or labyrinth is much more effective in attenuating photon radiation than neutrons. The design of a labyrinth for a photon source of known dose equivalent rate can be made using the albedo coefficients and the expression [1]:

$$H_{r_j} = \left( \frac{H_0}{d^2} \right) \left( \frac{\alpha_1 A_1}{d_1^2} \right) \prod_{i=2}^{i=j} \left( \frac{\alpha_i A_i}{d_i^2} \right) \quad (1)$$

Here  $H_{r_j}$  is the ambient dose equivalent rate after  $j$  sections of the maze (not counting the initial path to the wall  $d$ ) and  $\alpha$  are the albedo coefficients. Albedo coefficients can be found, *e.g.*, in Ref.[1]. They depend on the photon energy spectrum, the irradiated material, the angle of scattering and the orientation of the scattering surface. Typical values are in the range 0.1 to 0.001.  $A_1$  estimates the cross-sectional area of the wall struck by the initial beam,  $A_i$  is the cross-sectional area of the  $i$ -th leg of the labyrinth and the distances  $d_i$  are the radiation paths in the various legs of the labyrinth, and represent the inverse-square law dependence. After the first leg, it is usually a conservative assumption to use values for  $\alpha_i$  appropriate for 0.5 MeV photons. Equation (1) is probably most accurate if the ratio  $d_i^2/A_i$  lies between 4 and 40. The higher the incident photon energy the more conservative the formula is expected to be.

## References

- [1] NCRP Report No. 144 (2003)

### 8.4.7 Shielding Materials

Economic considerations play an increasingly important role in shielding design, among others, due to the size and complexity of many modern accelerators. Thus, detailed Monte Carlo simulations often allow the best optimization among the different constraints, such as radiation protection goals, structural properties, fire protection and financial limitations. Earth, concrete and iron are principally used for bulk shielding against stray radiation, especially neutrons, lead is very efficient to shield photons (both prompt photons and from radioactive decays) and marble has excellent low-activation properties [2].

**Earth** shielding has low cost and is efficient in attenuating low energy neutrons due to its hydrogen (water) content. In general, energy transfer to the shield in elastic scattering processes is most efficient if the masses of the interacting particles or nuclei are comparable which, in case of neutrons, is best achieved with hydrogen. Additionally, earth also contains elements with sufficiently high atomic number providing an efficient shielding against photons. It is “crackless”, *i.e.*, not prone to neutron leakage by “streaming” and has a density between 1.7 g/cm<sup>3</sup> and 2.2 g/cm<sup>3</sup>. Disadvantages are related to activation properties, such

as leaching of long-lived radionuclides (*e.g.*, tritium and <sup>22</sup>Na) into ground-water or the production of <sup>60</sup>Co, <sup>152</sup>Eu, <sup>154</sup>Eu and <sup>134</sup>Cs on trace elements. Table 1 gives the elemental composition of soil (global average, adapted from Ref.[1]). The water content (in percent of dry-weight) can vary between 0% and 30%.

Table 1: Elemental compositions of soil, concrete and heavy concrete (main components in dry-weight percent) [1]. The hydrogen content typically varies between 0.3% and 1.0% in concrete and can be higher in soil. The total does not add to 100%.

| Element | Dry soil | Concrete | Heavy concrete |
|---------|----------|----------|----------------|
| O       | 43.8     | 49.8     | 33.1           |
| Si      | 28.1     | 31.5     | 2.6            |
| Al      | 8.2      | 4.6      | 2.4            |
| Fe      | 5.1      | 1.3      | 47.4           |
| Ca      | 3.7      | 8.3      | 7.1            |
| Mg      | 2.1      | 0.26     | 0.93           |
| K       | 2.6      | 1.9      |                |
| Na      | 2.8      | 1.7      |                |
| Mn      | 0.07     |          | 0.2            |
| Ti      | 0.45     |          | 5.4            |
| S       |          |          | 0.14           |
| Cr      |          |          | 0.17           |
| V       |          |          | 0.31           |

**Concrete** is prepared from a mixture of about 13% cement, 7% water and 80% aggregate. It may be permanently poured in place or cast into movable blocks and has a density of about 2.35 g/cm<sup>3</sup>. Steel reinforcements can improve its structural strength. The use of a heavy material as part of the aggregate (*e.g.*, magnetite) increases its density (and, thus, reduces the required space) and effectiveness against photons. The water content, determining the neutron attenuation, decreases with time, sometimes to 50% of its initial value over a 20 year period. The elemental composition of ordinary concrete and an example for heavy concrete are given in Tab.1.

**Iron** combines a relatively high density (7.0–7.8 g/cm<sup>3</sup>) with a low cost to make it an attractive shielding material. While it is very efficient in attenuating high energy neutrons ( $E > 100$  MeV), neutrons below about 850 keV loose energy only by the inefficient process of elastic scattering. However, these low energy neutrons may be removed by placing a material containing light elements, such as concrete, outside the iron shield.

**Lead** has a high atomic number and is, therefore, frequently applied to shield photons, both prompt photons (*e.g.*, at electron accelerators or synchrotron radiation facilities) and photons emitted by radioactive decays. However, it has also a number of disadvantages, such as poor structural characteristics, a low melting point (327.4°C), a high chemical toxicity and poor activation properties, and its use should, thus, be justified.

**Marble** ( $\text{CaCO}_3$ ) is expensive but very efficient in reducing residual dose rates around highly activated components, such as dumps or collimators. It shows low activation and shields efficiently radioactive decay products of the components behind it.

## References

- [1] A.B. Chilton, J.K. Shultis, R.E. Faw, Principles of Radiation Shielding, Prentice Hall (1984)
- [2] R.H. Thomas in Handbook of Accelerator Physics and Engineering, World Scientific (2006) pp 670-672

### 8.4.8 Muons

Muons arise from the decay of pions and kaons, either in particle beams or in cascade induced by high energy hadrons. They can also be produced directly in high-energy hadron-nucleus interactions. Decay lengths of pions and kaons are 55.9 m and 7.51 m times the momentum (in  $\text{GeV}/c$ ) of the parent, respectively. Muons are weakly interacting particles and mainly lose energy by ionization, *i.e.*, they can only be stopped by “ranging them out”. Figure 1 shows range-energy curves for muons in various materials. Therefore, muon shielding is only important at accelerators above 10 GeV and is limited to the forward direction. At lower energy the shielding necessary to reduce radiation levels arising from nuclear cascade processes (typically a few metres of iron) is in excess of the ionization range of muons that could be relevant for radiation protection. At energies above 10 GeV the required shielding in the forward direction can conservatively be estimated by determining the mean range in the material from Fig.1 and adding about 20% to allow for range straggling.

In the energy range 10-30 GeV the muon fluence at a distance  $z$  behind the point of interaction of a proton of energy  $E$  (GeV) and where pions have a path length  $\Delta$  (m) in which to decay

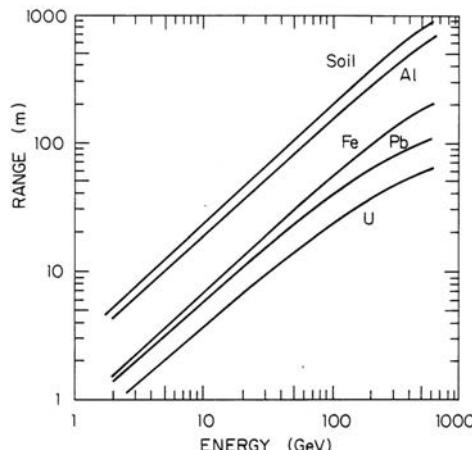


Figure 1: Range-energy curves for muons [1].

is given by

$$\phi(z) = 8.5 \times 10^{-2} E \Delta \exp(-\alpha t/E)/z^2. \quad (1)$$

Here,  $t$  is the shield thickness in meters and  $\alpha$  is an effective muon energy loss rate which has a value of 22 GeV/m in steel and 7.8 GeV/m in concrete.

For radiation protection calculations the fluence-to-dose equivalent conversion coefficient can be taken as 400 pSv cm<sup>2</sup> over the muon energy range 100 MeV - 100 GeV (see Fig.1 in Chapter 8.1.1).

## References

- [1] R.H. Thomas, G.R. Stevenson, IAEA Technical Report Series No. 283 (1988)

### 8.4.9 Skyshine

The prompt radiation field at a distance greater than a few hundred metres from an accelerator has two components, direct and scattered radiation. Often, the latter dominates, especially in situations where the roof shield of an accelerator enclosure is thinner than the side shield.

Skyshine consists mainly of low energy neutrons ( $E < 20$  MeV) originating from scattering processes in air of the radiation penetrating the roof shield. It depends strongly on intensity as well as angular and energy distribution of the neutrons exiting the shield and is, thus, most reliably assessed during accelerator design with a dedicated Monte Carlo simulation.

However, rough estimates can also be obtained from empirical parameterizations of dose equivalent  $H$  at a distance  $r$  from the accelerator [1]

$$H(r) = \frac{Q}{4\pi r^2} \exp(-r/\lambda). \quad (1)$$

This equation is valid at large distances ( $r > 60$  m) and considers the exponential attenuation of the skyshine neutrons in air with an effective attenuation length  $\lambda$ . The latter parameter increases with the average energy of the neutrons emitted from the shield but can also have larger values in case of multiple or extended sources. Typical values are between 150 m and 1000 m. The parameter  $Q$  represents several factors, such as the neutron emission rate, a neutron build up factor, the photon contribution associated with the neutron scattering processes as well as an average fluence-to-dose equivalent conversion coefficient.

## References

- [1] G.R. Stevenson and R.H. Thomas, *Health Phys.* 46 (1984) 115

## 8.5 RADIATION MEASUREMENTS

*S. Roesler, M. Silari, CERN*

The capacity to distinguish and measure the high-LET (mostly neutrons) and the low-LET components of the radiation field at workplaces is of primary importance to evaluate the exposure of personnel. At proton machines the dose equivalent outside a thick shield is mainly due to neutrons, with some contribution from photons and, to a minor extent, charged particles. At high-energy electron accelerators the dominant secondary radiation is made of high-energy neutrons, because the shielding is normally thick enough to absorb most of the bremsstrahlung photons. This section discusses some the methods employed for radiation monitoring at particle accelerator facilities. More complete information can be found in a recent review and references therein quoted [1]. The operating principles of the various types of radiation detectors are discussed in details in several textbooks (see, for example, Ref.[2]).

## References

- [1] P. Bilski et al., CERN Yellow Report CERN 2006-007 (2006)

- [2] G. F. Knoll, *Radiation detection and measurements*, John Wiley and sons

### 8.5.1 Neutron Dosimetry and Spectrometry

This section discusses the principal techniques employed for neutron monitoring: two are based on active instrumentation, moderator-type area monitors (rem counters) and Bonner spheres, and two employ passive dosimeters, bubble detectors and track-etched detectors. This discussion is restricted to those methods which show also a good response to high-energy neutrons, which contribute up to about 50% of the total ambient dose equivalent at high-energy accelerators.

#### 8.5.1.1 Rem counters

The type of instrument employed since many years for neutron monitoring is the so-called rem counter. A rem counter has a response function that approximately follows the curve of the conversion coefficients from neutron fluence to  $H^*(10)$  over a wide energy range. A commonly used instrument of this type is the Andersson-Braun (A-B) rem counter [1]. Its response is considered acceptable for neutron energies between thermal and approximately 10 MeV, although in reality the monitor underestimates  $H^*(10)$  in the energy range from thermal to about 1 eV and above a few MeV, and overestimates it in the interval 1 eV - 100 keV. The standard instrument consists of a thermal neutron detector enclosed within a moderator/attenuator assembly made up of an inner polyethylene moderator, a boron doped plastic attenuator (sometimes a cadmium layer is used instead) and an outer polyethylene moderator. The thermal neutron detector is usually either a  $\text{BF}_3$  or a  $^3\text{He}$  proportional counter. The moderators employed in the different versions available have either a cylindrical, cylindrical with a rounded edge, or spherical shape, the latter ensuring a more isotropic response.

Regardless of the construction details, the drop in sensitivity restricts the upper energy limit of a conventional rem counter to about 10 MeV. Above this value, the response falls sharply, leading to a drastic underestimation (about 40%) of the ambient dose equivalent, which increases with neutron energy. To overcome this limitation there exist versions of this instrument with response extended to high-energy neutrons. The original

device named LINUS was developed by Birattari *et al.* [2, 3], starting from a commercial A-B monitor by inserting a 1 cm thick lead layer between the boron doped plastic attenuator and the outer polyethylene moderator. Below 10 MeV, the response of the monitor is nearly the same as that of conventional moderator instruments. A number of models are now commercially available from various companies. The same concept of adding a lead liner has also been applied to Bonner sphere spectrometers (see next section).

### 8.5.1.2 Bonner sphere spectrometers

A multi-sphere spectrometer [4] (usually called Bonner Sphere Spectrometer, BSS) uses a thermal neutron detector at the center of moderating spheres of different diameters made of polyethylene. Fast neutrons are slowed down in the moderator and reach the detector as thermal ones, while the thermal neutrons initially present in the field are mostly captured in the moderator. Therefore the neutron energy, at which the sensitivity peaks, increases with sphere diameter. Due to the shape of the response functions, the energy resolution of the system is rather low but still satisfactory for radiation protection purposes.

A conventional BSS consisting of only polyethylene moderators has an inherent upper energy limit around 10-20 MeV and is unreliable in environments with a large contribution of high-energy neutrons, such as around intermediate and high-energy accelerators.

A BSS is normally employed with  $^6\text{Li}(\text{Eu})$  scintillators optically coupled to a photo-multiplier, or  $^3\text{He}$  or  $\text{BF}_3$  proportional counters. With  $^3\text{He}$  counters the discrimination with respect to gamma rays and noise is normally excellent. The  $^3\text{He}$  counters are fairly insensitive to radiations other than neutrons and their efficiency has proven to be stable with time. Also passive detectors are used in Bonner spheres, in order to measure in intense pulsed neutron fields or in cases where a low-intensity neutron field requires a very long integration time such as in some environmental measurements. Passive detectors include activation detectors sensitive to thermal neutrons, pairs of  $^6\text{Li}$  and  $^7\text{Li}$  fluoride thermo-luminescent detectors, and track detectors with radiators made of  $^{10}\text{B}$ ,  $^6\text{Li}$  or  $^{235}\text{U}$ .

The neutron spectrum is obtained by unfolding the experimental counts of the BSS with its response matrix by means of one of the several

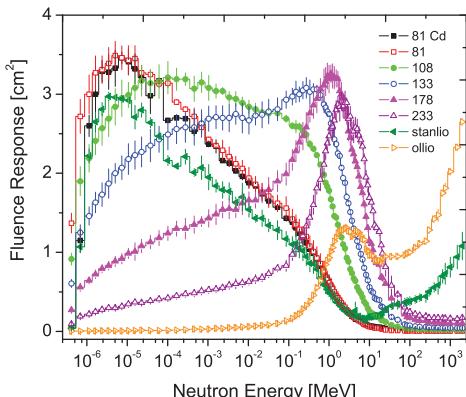


Figure 1: Absolute neutron fluence response of a BSS that includes two high-energy channels, labeled “stanolio” and “ollio” (from Ref.[11]).

codes developed over the years such as, *e.g.*, GRAVEL [5], MAXED [6], FRUIT [7], BONDI-97 [8] and BUNKIUT [9]. Use has also recently been made of unfolding techniques exploiting algorithms based on artificial intelligence [10]. The absolute fluence response of the BSS is expressed as the number of counts per incident neutron fluence. The response  $R(E)$  of each detector to direct neutrons with spectral fluence  $\Phi(E)$  is associated with the reading  $M$  of the detector via the expression:

$$M = \int R(E)\Phi(E) dE. \quad (1)$$

There are other contributions to the detector reading that have to be subtracted, either as background contributions or as “noise” to the real measurement.

In the past few years various groups have developed BSS with a response matrix extended to higher energies, following the same approach used in the development of the LINUS rem counter (see section 8.5.1.1), *i.e.*, by including a shell of high- $Z$  material in the moderator. The response functions of a BSS that includes two high-energy channels is shown in Figure 1.

### 8.5.1.3 Bubble detectors

“Super-heated emulsion” is the name adopted by the International Organization for Standardization (ISO) and the International Commission on Radiation Units and Measurements (ICRU) for detectors based on super-heated droplets suspended in

a gel, also known as bubble detectors or super-heated drop detectors [12]. The super-heated emulsion is contained in a vial and acts as a continuously sensitive, miniature bubble chamber. The total number of bubbles evolved from the radiation-induced nucleation of drops gives an integrated measure of the total neutron exposure. Various techniques exist to record and count the bubbles forming in super-heated emulsions, *e.g.*, visual inspection, automated reading with video cameras or acoustic counting. Further details on the physics and the operation of these detectors can be found in Refs.[12, 13, 14]. Super-heated emulsions are currently used as personal, area and environmental dosimeters, as well as neutron spectrometers. One of the advantages of bubble detectors is the possibility of determining an average ambient dose equivalent rate in a pulsed neutron field where active devices may suffer from dead-time losses or pulse pile-up. Another feature is that they are insensitive to low-LET radiation, X and  $\gamma$  rays as well as muons, which is a clear advantage when measuring the neutron component in mixed fields.

The  $H^*(10)$  response is underestimated for epithermal neutrons (up to about 100 keV), is fairly accurate in the neutron energy interval from 100 keV up to about 10 MeV, and is significantly underestimated for higher energies neutrons. Similarly with what has been achieved with rem counters and BSS, the sensitivity to high-energy neutrons (above 20 MeV) can be increased by encapsulating the dosimeter in a shell of high-Z material.

#### 8.5.1.4 Track etched detectors

Track etched detectors (TEDs) [15, 16] are based on the preferential dissolution of suitable, mostly insulator, materials along the damage “trails” of charged particles of sufficiently high-energy deposition density. The detectors are effectively not sensitive to radiation which deposits the energy through the interactions of particles with low LET. Neutron detection and dosimetry by means of TEDs can be performed through the registration of fission fragments incident on a TED from an adjacent radiator. In such cases inorganic TEDs (mica, glasses, minerals, *etc.*) can be used. Much more frequently polymer TED have been used which can register neutron induced lighter secondary charged particles, like protons, alpha-particles and other recoil nuclei. These secondary

particles can originate from nuclear reactions both in materials adjacent to a TED and those created inside the bulk of it. The TED response characteristics are generally sufficiently well known for neutrons with energies up to several hundred MeV [17]. These dosimeters are generally able to determine neutron ambient dose equivalent down to around 100  $\mu\text{Sv}$ .

## References

- [1] I.O. Andersson and J. Braun, Proc. of the IAEA Symposium on Neutron dosimetry, IAEA, Vienna, Vol. II (1963) 87
- [2] C. Birattari et al., Nucl. Instrum. Meth. A297 (1990) 250
- [3] C. Birattari et al., Radiat. Prot. Dosim. 76 (1998) 135
- [4] R. L. Bramblett et al., Nucl. Instrum. Meth. 9 (1960) 1
- [5] M. Matzke, Braunschweig, Germany, PTB, Bericht 19 (1994)
- [6] M. Reginatto, P. Goldhagen, Health Phys. 77 (1999) 579
- [7] R. Bedogni et al., Nucl. Instrum. Meth. A580 (2007) 1301
- [8] B. Mukherjee, Nucl. Instrum. Meth. A432 (1999) 305
- [9] K. A. Lowry, T. L. Johnson, Health Phys. 47 (1984) 587
- [10] H.R. Vega-Carrillo et al., Radiation Measurements 41 (2006) 425
- [11] C. Birattari et al., Nucl. Instrum. Meth. A620, (2010) 260
- [12] F. d'Errico, Nucl. Instrum. Meth. B184 (2001) 229
- [13] R.E. Apfel, Nucl. Instrum. Meth. 162 (1979) 603
- [14] H. Ing et al., Radiation Measurements 27 (1997) 1
- [15] R. L. Fleischer et al., Nuclear traces in solids, University of California Press, Berkley, 1975
- [16] K. Becker, Topics in Radiation Dosimetry, Ed. F. H. Attix, Academic Press, London, 1972, p. 79
- [17] L. Lindborg et al., EC Issue No. 140, Radiation Protection, ISBN 92-894-8448-9, EC 2004, Appendix A, p. 109

#### 8.5.2 Photon Dosimetry and Spectrometry

The most commonly used methods of gamma spectrometry with gamma-neutron discrimination in complex radiation fields is the use of NE213 or BC501A liquid scintillator cells coupled to a photomultiplier together with n- $\gamma$  discriminator based

on zero-crossing techniques (Z/C) [1, 2]. The n- $\gamma$  discrimination method is utilizing a difference in the intensity of the slow component of the light pulse in organic scintillators generated by recoil protons and electrons. The high-energy detection threshold depends of the size of the cell. A complementary spectrometric technique for n- $\gamma$  separation is the use of BGO semiconductor crystals. The detectors can be used up to photon energy of 30 MeV. This section will only discuss active measurement techniques.

### 8.5.2.1 BC501A liquid scintillator

The BC501A detector system allows measuring the energy distribution of neutrons and photons in the energy range from 1 MeV to about 100 MeV for neutrons and from 100 keV to 10 MeV for photons [3]. Fast neutrons and photons create, in the scintillation liquid cell, charged particles via elastic scattering and nuclear reactions and secondary electrons, respectively, with energy ranges between zero and the neutron incident energy or the Compton edge value for the photons. The secondary charged particles excite the scintillation molecules contained in the liquid. The decay of these molecules occurs through photon emission. Three decay modes associated to three decay times are known. The probability to excite each level depends on the nature of the secondary charged particle. The time characteristic of the scintillation light is then different for electrons and recoil protons, and therefore for photons and neutrons. This difference is used to discriminate the two types of radiation.

### 8.5.2.2 BGO photon spectrometer

In the detection of high-energy  $\gamma$ -rays ( $E_{\gamma} > 3$  MeV), the efficiency of semi-conductors (Ge) decreases rapidly with increasing  $E_{\gamma}$ . In this energy range scintillation crystals of bismuth germanate (BGO) or activated sodium iodide NaI(Tl) are more suitable as efficient detectors of gamma radiation. BGO scintillators offer an attractive alternative to NaI(Tl) because of their larger (typically by a factor of 2.5) total gamma absorption coefficient. Moreover, compared to NaI(Tl), BGO exhibits a much superior gamma-to-neutron detection ratio. Due to the bismuth high atomic number ( $Z = 83$ ) and high density ( $7.13 \text{ g/cm}^3$ ), the BGO is a very efficient gamma-ray absorber, such that it can provide spectral coverage from about 150 keV to 30 MeV. The use of a BGO

spectrometer in mixed neutron-photon fields requires, like for the NE213/BC501A spectrometer, taking into account the response due to photons induced by neutrons in the detector.

### 8.5.2.3 Geiger Müller counter

Geiger Müller (GM) counters are low cost devices, simple to operate. Since they only count radiation-induced events, any spectrometric information is lost. In general they are calibrated in terms of air kerma, for instance in a  $^{60}\text{Co}$  field, and then the air kerma and dose equivalent can be measured. The response of GM counters to photons is constant within 15% for energies up to 2 MeV and shows considerable energy dependence above. The neutron sensitivity of GM counter tubes is described by the  $k_u$  coefficient [4], which is used to correct the reading of the counter when used in a mixed neutron/photon field. The absolute spectral neutron fluence (or the relative spectral fluence and the neutron absorbed dose or kerma) must therefore be known to determine the number of neutron-induced counts, using the neutron fluence-to-kerma conversion coefficient for ICRU muscle tissue [5] and the energy dependent neutron sensitivity  $k_u$  coefficients.

## References

- [1] H. Klein, S. Neumann, Nucl. Instrum. Meth. A476 (2002) 132
- [2] A. Buffler et al., Nucl. Instrum. Meth. A476 (2002) 181
- [3] J. B. Czirr, Nucl. Instrum. Meth. 88 (1970) 321
- [4] ICRU Report 26, Bethesda, MD, USA (1977)
- [5] ICRU Report 63, Bethesda, MD, USA (2000)

### 8.5.3 Detectors Based on Micro-Dosimetric Principles

A separate class of detectors like tissue equivalent proportional counters (TEPC) and recombination chambers operate on micro-dosimetric principles. These devices allow the measurement of the total dose equivalent and to discriminate the low- and high-LET components of a radiation field.

TEPCs measure the probability distribution of absorbed dose  $D(y)$  in terms of lineal energy  $y$ , a stochastic quantity defined as the ratio of the energy imparted to the matter in a volume by a single-deposition event to the mean chord length

## Sec.8.6: INDUCED RADIOACTIVITY

in that volume [1]. TEPCs are proportional counters containing a tissue-equivalent (TE) gas inside a TE plastic cavity chamber. Acting on the gas pressure, it is possible to simulate the energy deposition events in microscopic volumes. From the probability distribution of absorbed dose  $D(y)$  one can evaluate the dose equivalent through a function  $Q(y)$  [2] which relates the quality factor to the linear energy because  $y$  can be used as an approximation of the linear energy transfer, LET.

Recombination chambers, *i.e.*, high-pressure ionization chambers involving the initial recombination of ions, are usually considered as detectors of the total ambient dose equivalent  $H^*(10)$ , or, separately, the ambient absorbed dose and average quality factor of a mixed radiation field.

Another type of recombination chamber can be applied for photon dosimetry in mixed photons/neutrons fields using the two-detector technique. This is a high-pressure, hydrogen-free chamber, usually with graphite or aluminum electrodes and filled with carbon dioxide up to about 3 MPa. The chamber is operating at low polarizing voltage, in conditions of strong initial recombination of ions in high-LET particles tracks. The neutron-to-photon sensitivity ratio of the chamber is strongly reduced by the initial recombination of ions created in the tracks of alpha particles and nuclear recoils generated by neutrons. The photon kerma can be determined separately, using this method, if the energy of accompanying neutrons does not exceed some tens of MeV [3].

## References

- [1] ICRU Report 36, Bethesda, MD, USA (1983)
- [2] ICRU Report 40, Bethesda, MD, USA (1986)
- [3] N. Golnik et al., Radiat. Prot. Dosim. 88 (2000) 135

## 8.6 INDUCED RADIOACTIVITY

*S. Roesler, M. Silari, CERN*

Neutrons are not affected by the Coulomb barrier of nuclei and can, thus, react at any energy and produce radioactive nuclides. Neutron capture dominates for thermal neutrons, while with increasing energy reactions of type (n,p), (n, $\alpha$ ), (n,2n), *etc.* occur. High-energy neutrons cause spallation reactions that can produce any nuclide lighter than the target nucleus.

Charged particles with energies lower than the Coulomb barrier (few MeV) do not effectively react with nuclei. As soon as the energy exceeds the Coulomb barrier compound nuclei might be formed which de-excite by the emission of photons, nucleons or light nuclei (*e.g.*, for protons, reactions of type (p,n), (p,d), (p, $\alpha$ ), *etc.*). Similar to neutrons, high-energy charged particles interact in spallation reactions.

Also electromagnetic particles may cause activation through photo-nuclear interactions, although with a much lower cross section than hadronic reactions (at high energy lower by the fine structure constant). Thus, activation by electrons and photons is typically not a concern at hadron accelerators while it might be important at electron accelerators. Threshold energies for photo-nuclear reactions are a few MeV, depending on the target material. Just above threshold, so-called Giant Dipole resonance reactions dominate in which the nucleus de-excites through the emission of neutrons, protons and light nuclei.

### 8.6.1 Fundamental Principles

Radioactive decay is a random process characterized by a decay constant  $\lambda$ . If a total of  $N_{\text{tot}}(t)$  atoms of a radionuclide are present at time  $t$ , the total activity  $A_{\text{tot}}(t)$  is determined by

$$A_{\text{tot}}(t) = -\frac{dN_{\text{tot}}(t)}{dt} = \lambda N_{\text{tot}}(t) \quad (1)$$

with the solution at  $t = T$

$$A_{\text{tot}}(T) = A_{\text{tot}}(0) e^{-\lambda T}. \quad (2)$$

Often, the time required to decay half of the original activity, the half-life,  $t_{1/2}$ , is given and is related to the decay constant by

$$t_{1/2} = \frac{\ln 2}{\lambda}. \quad (3)$$

Assuming a steady irradiation of some material by a spatially uniform fluence rate,  $\phi$  ( $\text{cm}^{-2}\text{s}^{-1}$ ), the density of atoms,  $n(t)$ , of the radionuclide of interest per unit volume at time  $t$  ( $\text{cm}^{-3}$ ) during that irradiation will be governed by

$$\frac{dn(t)}{dt} = -\lambda n(t) + N\sigma\phi \quad (4)$$

where  $\sigma$  is the production cross section ( $\text{cm}^2$ ) and  $N$  is the density of target atoms ( $\text{cm}^{-3}$ ). The equation has the solution

$$n(t) = \frac{N\sigma\phi}{\lambda} \left(1 - e^{-\lambda t}\right) \quad (5)$$

where the specific activity during irradiation is given by  $a(t) = \lambda n(t)$ . For  $t \gg t_{1/2}$  Eq.(5) yields  $a(t) = a_{\text{sat}} = N\sigma\phi$ , i.e., the saturation activity equals the production rate.

### 8.6.2 Solid Materials

The most important medium- and long-lived radionuclides produced in typical accelerator materials are given in Tab.1. As can be seen, the heavier the elements are in the respective material the more radionuclides can be created. Thus, light material should be preferred, if possible, in the construction of accelerator components. For example, aluminum supports have better radiological characteristics than steel supports due to the significantly lower number of produced nuclides.

Reactions on trace elements in the materials give rise to additional nuclides which might also be important, especially if they are long-lived. A typical example is  $^{60}\text{Co}$  produced by thermal neutron capture reactions on traces of cobalt in aluminum or iron components. This nuclide can dominate the activity in the respective component many years after irradiation when most other nuclides have already decayed.

Activation properties of materials used for accelerator construction must be considered already during the design as they may have direct impact on later handling (maintenance, repair) and waste disposal. Gamma-emitting nuclides dominate residual dose rates at longer decay times (more than one day) while at short decay times  $\beta^+$  emitters are also important (through photons produced by  $\beta^+$  annihilation). Due to their short range,  $\beta^-$  emitters are relevant, for example, only for dose to the skin and eyes or for doses due to inhalation or ingestion.

Figures 1 and 2 show the contributions of gamma and  $\beta^+$  emitters, respectively, to the total dose rate close to an activated copper sample [1]. Typically, dose rates at a certain decay time are mainly determined by radionuclides having a half-life of the order of the decay time. Extended irradiation periods might be an exception to this general rule as in this case the activity of long-lived nuclides can built up sufficiently so

Table 1: Nuclides of radiological importance in elements of typical accelerator materials. The last column indicates the respective half-lives.

| Element or material | Nuclide           | $t_{1/2}$ |
|---------------------|-------------------|-----------|
| Carbon              | $^3\text{H}$      | 12.3 a    |
|                     | $^7\text{Be}$     | 53.29d    |
|                     | $^{11}\text{C}$   | 20.38m    |
| Aluminum            | all above plus    |           |
|                     | $^{22}\text{Na}$  | 2.6a      |
| Iron                | $^{24}\text{Na}$  | 15.0h     |
|                     | all above plus    |           |
|                     | $^{m44}\text{Sc}$ | 2.44d     |
|                     | $^{46}\text{Sc}$  | 83.8d     |
|                     | $^{48}\text{Sc}$  | 1.81d     |
|                     | $^{48}\text{V}$   | 16.0d     |
|                     | $^{51}\text{Cr}$  | 27.70d    |
|                     | $^{52}\text{Mn}$  | 5.6d      |
|                     | $^{54}\text{Mn}$  | 312.1d    |
|                     | $^{55}\text{Fe}$  | 2.73a     |
|                     | $^{59}\text{Fe}$  | 44.50d    |
|                     | $^{55}\text{Co}$  | 17.54h    |
|                     | $^{56}\text{Co}$  | 77.3d     |
|                     | $^{57}\text{Co}$  | 271.8d    |
|                     | $^{58}\text{Co}$  | 70.82d    |
|                     | Stainless Steel   |           |
|                     | all above plus    |           |
|                     | $^{60}\text{Co}$  | 5.27y     |
|                     | $^{57}\text{Ni}$  | 35.6h     |
| Copper              | all above plus    |           |
|                     | $^{63}\text{Ni}$  | 100a      |
|                     | $^{61}\text{Cu}$  | 3.4h      |
|                     | $^{64}\text{Cu}$  | 12.7h     |
|                     | $^{65}\text{Zn}$  | 244.3d    |

that it dominates that one of short-lived even at short cooling times.

Activation in concrete is dominated by  $^{24}\text{Na}$  (short decay times) and  $^{22}\text{Na}$  (long decay times). Both nuclides can be produced either by low-energy neutron reactions on the sodium-component in the concrete or by spallation reactions on silicon and calcium. At long decay times nuclides of radiological interest in activated concrete can also be  $^{60}\text{Co}$ ,  $^{152}\text{Eu}$ ,  $^{154}\text{Eu}$  and  $^{134}\text{Cs}$ , all of which produced by  $(n,\gamma)$ -reactions with traces of natural cobalt, europium and cesium. Thus, such trace elements might be important even if their content in concrete is only a few parts per million or less by weight.

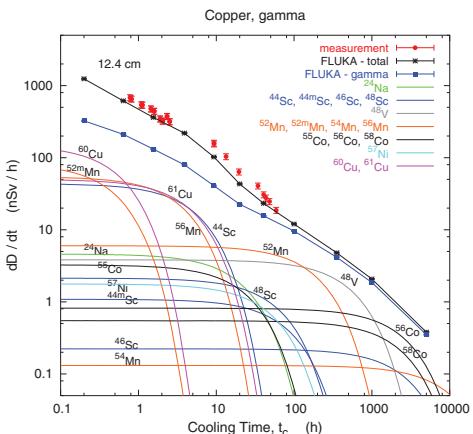


Figure 1: Contribution of individual gamma-emitting nuclides to the total dose rate at 12.4 cm distance to an activated copper sample [1].

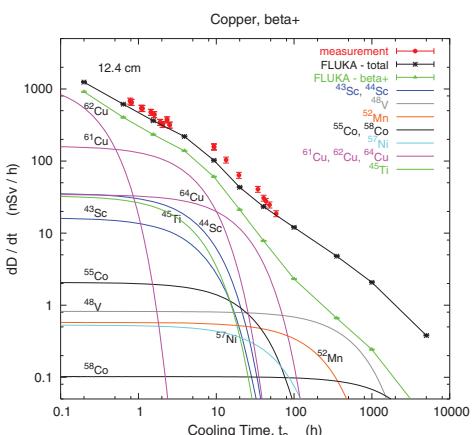


Figure 2: Contribution of individual positron-emitting nuclides to the total dose rate at 12.4 cm distance to an activated copper sample [1].

The explicit simulation of radionuclide production with general-purpose Monte Carlo codes has become the most commonly applied method to calculate induced radioactivity and its radiological consequences. Nevertheless, other more approximative approaches, such as “ $\omega$ -factors” [2], can still be useful for fast order-of-magnitude estimates. These  $\omega$ -factors give the dose rate per unit star density (inelastic reactions above a certain energy threshold, e.g., 50 MeV) in contact to an extended, uniformly activated object after a 30-day

irradiation and 1-day decay. The  $\omega$  factor for steel or iron is  $\simeq 3 \times 10^{-12}$  (Sv cm<sup>3</sup>/star). This does not include possible contributions from thermal-neutron activation.

These phenomena can also be used as "activation detectors" for neutron dosimetry in accelerator environments [3].

## References

- [1] S. Roesler et al, Proc. of the Sixth International Meeting on Nuclear Applications of Accelerator Technology (2003) 655
  - [2] R.H. Thomas, G.R. Stevenson, IAEA Technical Report Series No. 283 (1988)
  - [3] R.H. Thomas in Handbook of Accelerator Physics and Technology, World Scientific (2006) p. 675

### 8.6.3 Liquids

At accelerators, liquids are mainly used for cooling purposes (*e.g.*, demineralized water, liquid helium), but also liquid targets exist (*e.g.*, liquid mercury).

Spallation reactions of secondary particle showers on oxygen in demineralized water can create tritium ( $t_{1/2}=12.3$ a),  ${}^7\text{Be}$  ( $t_{1/2}=53.29$ d) and a number of short-lived  $\beta^+$ -emitters ( ${}^{11}\text{C}$ ,  ${}^{13}\text{N}$ ,  ${}^{15}\text{O}$ ). The production of tritium by thermal neutron capture on natural hydrogen can be neglected in most application due to the low abundance of deuterons and small cross section. Sometimes cooling water circuits also contain nuclides from corrosion products (e.g., cobalt nuclides) of which, however, a large fraction is collected, together with  ${}^7\text{Be}$ , in the resin of ion exchanger cartridges. In natural water, radionuclides can also be produced in reactions with trace elements (minerals).

During accelerator design, the activation of cooling liquids is most conveniently assessed by folding fluence spectra with energy-dependent nuclide production cross sections. Also a direct calculation is possible with Monte Carlo codes for nuclides produced on oxygen, while this direct method would fail for nuclides from trace elements due to lack of statistical significance.

Activated cooling liquids pose contamination hazards during interventions on accelerator components but may also cause external irradiation close to pipes and cartridges. Although tritium decay proceeds only via emission of a low energy electron, its concentration in water, especially if

released off-site, has become a critical parameter as it may attract the attention of the public.

### 8.6.4 Air

Airborne radionuclides are produced mainly in interactions of beam particles or associated showers of secondary particles with air molecules. Other sources might be activated dust or out-gassing of nuclides from activated accelerator components. The latter two sources, however, are typically of lower importance and can only be assessed by measurements.

Table 1 gives the nuclides of highest radiological importance. At hadron and ion accelerators, most of them are created by spallation reactions on air molecules. Only  $^{41}\text{Ar}$  results from thermal neutron capture reactions on argon ( $\sigma_{\text{th}}=660 \text{ mb}$ ). At electron accelerators, photonuclear interactions of type ( $\gamma, n$ ) contribute to the production of  $^{13}\text{N}$  and  $^{15}\text{O}$ . Although the radiological impact of  $^3\text{H}$  in air is small it easily attaches to humidity and can reach waste water circuits, especially through condensation on air-conditioning units.

Apart from the list in Tab.1 specific situations and exposure pathways may require the consideration of further nuclides, such as  $^{32}\text{P}$  ( $t_{1/2}=14.26\text{d}$ ) which is produced by spallation reactions on argon. This nuclide can reach the milk consumed by infants through ground deposition on grazing land and, thus, dominate the committed dose due to ingestion.

The low density of air usually renders a direct calculation of air activation by Monte Carlo models highly inefficient. Instead, particle fluence spectra are multiplied with energy-dependent nuclide production cross sections, obtained from Monte Carlo models, experimental data or both (so-called evaluated cross sections). This yields

Table 1: Airborne nuclides of radiological importance. The second column indicates the respective half-lives.

| Nuclide          | $t_{1/2}$ |
|------------------|-----------|
| $^3\text{H}$     | 12.3 a    |
| $^7\text{Be}$    | 53.29d    |
| $^{11}\text{C}$  | 20.38m    |
| $^{13}\text{N}$  | 9.96m     |
| $^{15}\text{O}$  | 2.03m     |
| $^{41}\text{Ar}$ | 1.83h     |

nuclide production rates per unit volume or, after application of Eq.5 in Chapter 8.6.1, specific activity.

Results of air activation studies play a crucial role in the design of the ventilation system of an accelerator. Closed circuits, that are flushed with fresh air only prior to access, have the advantage of reducing the total annual release of short-lived nuclides. However, the concentration of long-lived nuclides may build up and lead to undue exposure if released at once in a short time period which would not benefit from changing wind-conditions. In addition, tritium can build up, attach to water and accumulate, e.g., in sumps. On the other hand, constant venting with fresh causes an increased annual release of short-lived nuclides while profiting from natural dilution of long-lived ones. Apart from environmental aspects, ventilation systems have safety functions in ensuring the containment of radioactive gases and should follow international standards [1].

Adjustments for the presence of a ventilation can be made introducing an effective decay constant,  $\lambda'$ , that includes the physical decay constant along with a ventilation term

$$\lambda' = \lambda + \frac{D}{V} \quad (1)$$

with  $D$  being the ventilation rate (volume of air exchanged per unit time) and  $V$  the enclosure volume. Thus, with ventilation the saturation activity,  $a'_{\text{sat}}$ , becomes

$$a'_{\text{sat}} = \frac{\lambda a_{\text{sat}}}{\lambda + D/V}. \quad (2)$$

## References

- [1] International Organization for Standardization  
ISO 17873, International Standard (2004),  
<http://www.iso.org>

## 8.7 RADIATION DAMAGE THRESHOLDS

*H. Schönbacher, CERN*

### 8.7.1 Organic Materials

Apart from electronic and optical devices, the organic materials are the ones most sensitive to radiation. As a consequence of this, a large number of radiation tests have been made on these materials and the results are extensively documented [1]-[15].

Some characteristic physical, mechanical and electrical properties of thermoplastic and thermosetting materials are summarized in Tab.1. These values are only a general indication since they depend on numerous parameters, such as the composition and quantity of the base resin, the hardener, the accelerator, the filler, and other additives, as well as on the curing conditions, etc.

**Test methods** In selecting and classifying materials according to their radiation resistance, not all of the properties listed in Table 1 could be tested. According to the recommendations of the International Electrotechnical Commission (IEC), the most radiation-sensitive property is chosen as the reference critical property. For the flexible thermoplastic materials this critical property is the elongation at break and for the thermosetting resins and composites it is the flexural strength. The end-point criterion is 50% of the initial value of the critical property [16].

**Radiation index RI** RI is defined in IEC 544-4 as the  $\log_{10}$  of the absorbed dose in grays (rounded down to two significant digits) at which the critical property is reduced to 50% of its initial value, under specified conditions of irradiation and tests.

**Environment and long term aging effects** Environmental conditions are very important for the assessment of radiation damage and shall be well-controlled and documented during the measurement of radiation effects in polymers. Environmental parameters include temperature, reactive medium, and mechanical and electrical stresses present during the irradiation. If air is present, radiation-induced species can enter into reactions with oxygen that would not occur in its absence. This is responsible for an observed influence of the absorbed dose rate for certain types of polymers if irradiated in air. As a result, the resistance may be as much as orders of magnitude lower than when the sample is irradiated under vacuum or in the presence of inert gas. This is generally called the “dose-rate effect”, which is described and reviewed in references [6]-[11].

The irradiation time can become relevant because of time-dependent complications caused by: a) physical effects such as diffusion-limited oxidation [17], and b) chemical phenomena such as rate-determining hydroperoxide breakdown reactions [18].

The typical diffusion-limited effects are commonly observed in radiation studies of polymers

in air. Their importance depends upon the interrelationship of the geometry of the polymer with the oxygen permeation and consumption rates, both of which depend upon temperature [17]. This means that the irradiation of thick samples in air may result in oxidation only near the air-exposed surfaces of the sample, resulting in material property changes similar to those obtained by irradiation in an oxygen-free environment. Therefore, when the material is to be used in air for a long period of time at a low dose rate, depositing the same total dose at a high dose rate in a short exposure period may not determine its durability.

Radiation-induced reactions will be influenced by *temperature*. An increase in reaction rate with temperature can result in a synergistic effect of radiation and heat. In the case of the more commonly used thermal aging prediction the Arrhenius method is employed; this makes use of an equation based on fundamental chemical kinetics. Despite considerable ongoing investigations of radiation aging methodologies, this field is not well developed. It should be noted that sequential application of radiation and heat, as it is frequently practiced, can give very different results depending on the order in which they are performed [18].

### Radiation damage tests

At ambient temperature In Tab.2, a classification of the materials in decreasing order of radiation resistance is given for rubber, thermoplastic and thermosetting materials respectively. These appreciations can only serve as a general guideline; environmental conditions such as temperature, humidity and dose rate, as well as additives, influence the radiation behavior of materials.

Detailed results are presented in [13] to [15] in the form of tables and graphs. The most common rubber and thermoplastic materials, of which data are presented there, are: ethylacrylate rubber (EAR); ethylene-propylene diene monomer rubber (EPDM); ethylene-propylene rubbers (EPR); silicone rubber (SIR); polychloroprene rubber (PCR); acrylonitrile rubber (ACR) ethylene vinyl acetate copolymers (EVA); linear or cross-linked polyethylenes (PE and XLPE) polyolefins; polyurethanes (PUR).

The most common thermosetting materials that are dealt with are: epoxy resins (EP); polyester resins (PLE); poly-urethane resins (PUR); aromatic cured epoxy (ACE); silicone (SI); melamine-formaldehyde (MF); urea-formaldehyde (UF); aniline-formaldehyde (AF).

Table 1 Properties of Some Thermoset and Thermoplastic Materials

| Mat'l         | $\rho$ | $H_2O_{abs}$ | $\kappa$ | $\alpha_{AT}$ | TS    | $\Delta L$ | E     | FS     | IS     | $\rho_v$            | DS    | Itg     | $\epsilon_r$ | HDT  |
|---------------|--------|--------------|----------|---------------|-------|------------|-------|--------|--------|---------------------|-------|---------|--------------|------|
| Epoxy         |        |              |          |               |       |            |       |        |        |                     |       |         |              |      |
| Bisph. A      | 1.15   | .1-.2        | .2-.5    | 6             | 70-80 | 4.4        | 3.5   | 80-130 | 1-2.7  | 6.1                 | >16   | .032    | 3.4          | 110  |
| Glass fill    | 2-2.1  | .02-.8       | 1.2      | .6            | ~375  | -          | 30    | 360    | 6.4-8  | -                   | 18-22 | .024    | ~4.7         | -    |
| Mineral       | 1.8-2  | .3-.8        | -        | -             | 70    | -          | 10-15 | ~160   | 2-2.7  | 1.5                 | 15-16 | .013    | 4-4.6        | -    |
| Novolac       | 1.2    | -            | -        | 3             | 70    | 2-5        | 3.5   | 60-100 | 2.7    | .21                 | -     | .029    | 3.5          | 150  |
| Glass fill    | 1.97   | ~.05         | -        | -             | ~375  | -          | 21-22 | 390    | 7-9    | -                   | -     | .015    | 5.1          | -    |
| Mineral       | 1.7    | .11-.2       | -        | -             | 38    | -          | -     | 70-80  | ~2.5   | .14-.5              | 12-16 | -       | ~4.5         | -    |
| PEEK          | 1.32   | .06-.2       | -        | 5             | 95    | >25        | 3.6   | 180    | -      | 10                  | 30    | .003    | 3.3          | 340  |
| 30% Glass     | 1.48   | .14          | -        | 2.5           | 170   | 2          | 9.7   | -      | -      | -                   | -     | -       | -            | 340  |
| Phenolic      | 1.3    | .3-.4        | -        | 4.3           | 14-63 | ~5         | 1.4-3 | 84-100 | 2.7-4  | <10 <sup>-3</sup>   | 10-16 | .04-    | 4-9.7        | 150  |
| Fiber, fabric | ~1.3   | .5-1.6       | .3       | 3             | 31-63 | ~5         | 6-10  | 49-110 | 2-36   | <10 <sup>-3</sup>   | 4-7   | .03-.08 | 5-7          | 250  |
| Polyimide     | 1.47   | .68          | .7       | 5.4           | 74    | <1.5       | 3.2   | 100    | 5      | 10-10 <sup>2</sup>  | 22    | .005    | 3.4          | 300  |
| Glass fill    | 1.9    | .2           | -        | 1.5           | 210   | <1         | ~29   | 350    | 17     | 9.2                 | -     | .0055   | 4.7          | 350  |
| Polyester     | ~1.3   | .03-.4       | .17      | 7             | 35-81 | 1.7-2.6    | 3-4.6 | 45-91  | 1.6-10 | .3-2                | 10-17 | .01-.03 | 3-4          | >50  |
| Glass fill    | 1.2-2  | .1-2         | 2.-2.6   | 2             | 42-90 | .3-5       | 4-12  | 84-150 | 40-54  | 10 <sup>-3</sup> -1 | 6-14  | 1.1-.04 | 4.5-6        | 200  |
| Polyurethane  | 1.21   | .3-.9        | -        | -             | 45-60 | -          | 3.3-8 | -      | >5.4   | <0.1                | 20    | .03-.05 | 3.3-4        | -    |
| Silicone      |        |              |          |               |       |            |       |        |        |                     |       |         |              |      |
| Glass filled  | 1.88   | .1-.3        | .3       | 6             | 28-56 | <.3        | 15-17 | 91-133 | 50     | .3                  | 10-11 | <.02    | 4.4          | >450 |
| Mineral       | 2-2.8  | .05-.2       | .6       | 5             | 17-31 | <.3        | 9-16  | 49-70  | 2      | .1                  | 11-16 | <.01    | 3.4-5        | >270 |

Table 2 Radiation tolerance of organic materials. Entries =  $\log_{10}$  of dose, [ gray] for no compromise; (xx) some compromise but useable.

|              |           |               |           |           |           |                  |                |                   |               |           |
|--------------|-----------|---------------|-----------|-----------|-----------|------------------|----------------|-------------------|---------------|-----------|
| PUR          | 5.4 (6.8) | Liq. Xtal LCP | 7 (7.8)   | PSU       | 6 (6.4)   | PMMA             | 4 (5)          | PUR               | 6.4 (7.7)     |           |
| EPR/EPDM     | 5.4 (6.4) | PEI           | 7 (7.5)   | Polyamide | 4.6       | 5.7 (6.7)        | POM            | 3.7 (4.6)         | PLE/ mineral  | 6.2 (7.4) |
| SBR          | 5.3 (6.4) | PAI           | 7 (7.5)   | PPO       | 5.7 (6.7) | PTFE             | 2.5 (4)        | SI/ unfilled      | 6 (7)         |           |
| EVA          | 5.2 (5.9) | PPS           | 6.7 (7.8) | ABS       | 5.7 (6.7) | Epoxy glass      | 7.8 (8)        | EP                | 5 (6.4)       |           |
| PCR          | 5.2 (5.9) | PEEK          | 6.4 (7.4) | PE        | 5 (6)     | Phenolic/ glass  | 7.8 (8)        | Phenolic/ no fill | 4.5 (6)       |           |
| ACR          | 5.2 (5.9) | PS            | 6.4 (7.4) | PETP      | 4.6 (6)   | " / mineral      | 7.6 (8)        | MF                | 4.7 (5.7)     |           |
| EAR/EEA      | 4.8 (5.8) | PI+siloxane   | 6.1 (6.7) | PC        | 4.6 (6)   | ACE special      | 7.3 (8)        | UF                | 4.5 (5.4)     |           |
| SIR          | 4.5 (5.5) | PAR           | 6.1 (6.7) | Polyamide | 6.6       | 4.2 (5.4)        | SI/ glass fill | 7.1 (7.7)         | PLE/ unfilled | 4 (5)     |
| Butyl rubber | 4 (4.5)   | PAA           | 6 (6.7)   | CA        | 4.2 (5.2) | SI/ mineral fill | 7.1 (7.7)      | AF                | 3.6 (5)       |           |
| PI           | 7 (7.8)   | PES           | 6 (6.4)   | PP        | 4.2 (5)   | PLE/ glass fill  | 7.1 (7.7)      |                   |               |           |

Further, rigid and high performance thermoplastics are presented: polyimide (PI); liquid crystal polymer (LCP); poly-ether-imide (PEI); poly-amide-imide (PAI); poly-phenyl-sulfide (PPS); poly-ether-ketone (PEEK); polystyrene (PS); polyacrylate (PAR); poly-aryl-amide (PAA); poly-ether-sulfide (PES); polysulfone (PSU); poly-phenol-oxide (PPO); acrylonitrile-butadiene-styrene (ABS), poly-ethylene (PE); poly-ethylene-terephthalate (PETP); polycarbonate (PC); polyamide (PA), cellulose acetate (CA); polypropylene (PP); polymethylmethacrylate (PMMA); poly-oxy-methylene (POM); poly-tetra-fluoro-ethylene (PTFE)

In the individual pages of results in [13, 14], the description of the material and the name of the supplier if available, the mean values (and the standard deviation) of the mechanical properties are given. These are, for flexible plastics tensile strength, elongation and Shore hardness. For rigid plastics these are the flexural strength, deformation at break and flexural modulus. They appear in a table and graph together with the absorbed doses and the corresponding dose rates. The RI for the corresponding dose rate is also given. Fig.1 displays a typical, but reduced in size, data page from [13, 14]

At cryogenic temperatures With the next generation of high-energy particle accelerators and detectors, many materials will have to be used at temperatures as low as 2 K. It is well known that thermal and mechanical properties of polymers are temperature sensitive; for example, the specific heat generally falls by more than two orders of magnitude between 300 K and 4K, and the ultimate deformation is reduced to a few percent. It is therefore necessary to check whether the mechanical radiation tests performed at room temperature (RT) are still representative of the radiation degradation at cryogenic temperature.

From data published in [20] one can see that significant differences exist between material types and, to a lesser extent, within the same material type depending on the irradiation temperature.

Common rubber and polyolefin cable-insulating materials are not suitable for low-temperature applications. Even without irradiation their elongation at break is reduced to less than 3%.

The radiation degradation of thermoplastic materials is less pronounced at low temperature.

|           |                                |              |
|-----------|--------------------------------|--------------|
| Material: | Prepreg                        | TIS No. R538 |
| Type      | Vetronite epoxy G11            |              |
| Supplier  | von Roll Isola                 |              |
| Remarks:  | proposed LHC magnet insulation |              |

#### Radiation Test Results (IEC 544; ISO 178)

| Dose<br>[MGy] | FS<br>[MPa]        | $\Delta l$<br>%  | E<br>[GPa]             |
|---------------|--------------------|------------------|------------------------|
| 0             | 557±23<br>(986±2)* | 2.5±0.1<br>(7.2) | 29.8±8.8<br>(28.6±0.2) |
| 5             | 484±16             | 2.2±0.2          | 25.7±0.9               |
| 10            | 448±24             | 2.0±0.1          | 26.6±0.6               |
| 14            | (497±52)           | (2.6±0.2)        | (19.8±0.8)             |
| 37            | (226±32)           | (1.4±0.2)        | (15.4±1.3)             |
| 50            | 295±22             | 1.4±0.1          | 26.1±0.5               |
| 100           | 211±20             | 1.1±0.1          | 22.5±1.3               |
| RI            | 7.7(7.2)           | 7.8(7)           | 7(~7.6)                |

\* entries in ( ) = data taken at 77 K

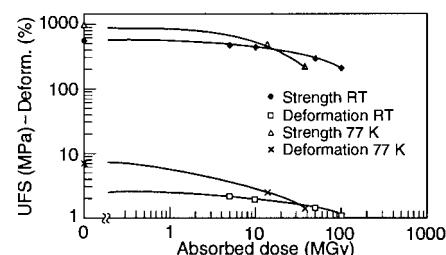


Figure 1: Reduced typical data page.

This could be related to the absence of oxygen during irradiation. The effect is very pronounced with thin films and makes most of the high-performance thermoplastics (including Mylar and Kapton films [20]) suitable for use in accelerator magnet cryostats.

No significant influence of the irradiation temperature is observed on the radiation degradation of thermosets and composites (see Fig.1 from [20]). Therefore, good indications for the radiation response of these types of materials for an application in a cryogenic environment can be assessed from RT tests after the initial properties have been measured at low temperature.

More information and specific test data can be found in the cited literature [19]-[28].

#### References

- [1] M.H. Van de Voorde, CERN 70-5-(1970)
- [2] M.H. Van de Voorde, C. Restat, CERN 72-7 (1972)

- [3] D.C. Phillips et al, CERN 81-05 (1981)
- [4] R.I. Keiser, M. Mottier, CERN 82-05 (1982)
- [5] G. Liptak et al, CERN 85-02 (1985)
- [6] R.L. Clough, Radiation resistant polymers, in Encyclopedia of polymer Science and Engineering, Wiley (1988), 20 (13 2nd ed., 1988, p.667).
- [7] Proc. Research Coordination Meeting on Radiation Damage to Organic Materials in Nuclear Reactor and Radiation Environment (1989), (IAEA Tech. Rep. Ser., Vienna, 1990)
- [8] IAEA-TECDOC-551: Radiation damage to organic materials in nuclear reactors and radiation environments (1990)
- [9] D.W. Clegg, A.A. Collyer, Irradiation effects on polymers. Elsevier Appl. Sc. (1991)
- [10] R.L. Clough, S.W. Shalaby, Radiation effects on polymers, ACS Symposium Series 475 (1991)
- [11] R.L. Clough, S.W. Shalaby, Irradiation of polymers, ACS Symposium Series No.620 (1996)
- [12] Proc. Int. Workshop on Advanced Materials for High Precision Detectors (1994), CERN 94-07, Part IV.
- [13] H. Schönbacher, M. Tavlet, CERN 89-12 (1989)
- [14] H. Schönbacher, A. Stolarz-Izicka, CERN 79-08 (1979), 2nd ed. in preparation
- [15] P. Beynel, P. Maier, H. Schönbacher, CERN 82-10 (1982)
- [16] Int. Electrotechnical Commission, Geneva, Publication No. 554, Part I: IEC 554-1 (1994); Part II: IEC 554-2 (1994); Part IV: IEC 554-4 (1985)
- [17] Int. Electrotechnical Commission, Geneva, Technical Report No.1244, Part 1: IEC 1244-1 (1993); Part 2: IEC 1244-2 (1996); Part 3
- [18] R.L. Clough, K.J. Gillen, Jour. Polym. Sci., Polym. Chem. Ed. 19 (8) 2041 (1981)
- [19] M. Van de Voorde, CERN 77-03 (1977)
- [20] K. Humer et al, CERN 96-05 (1996)
- [21] N.A. Munshi, H.W. Weber, Adv. Cryog. Eng. Mast., 38, 233-239 (1992); E.K. Tschegg, K. Humer, H.W. Weber, J. Mat. Sci. 28 (1993) 2471
- [22] E.K. Tschegg, K. Humer, H.W. Weber, J. Mat. Sci., 30 (1995) 1251
- [23] K. Humer et al, J. Nucl. Mat. 212-215 (1994) 849
- [24] K. Humer et al, Proc. 18th Symposium on Fusion Technology (1994); Fusion Technology 2 (1995) 973
- [25] S. Spiessberger et al, Adv. Cryog. Eng. Mat., Vol. 42
- [26] K. Humer, H.W. Weber, E.K. Tschegg, Cryogenics 35 (1995) 871
- [27] A. Spindel, SSCL-635 (1993)
- [28] K. Humer et al, Int. Cryogenic Materials Conf., Advances in Cryogenic Engineering Materials, Vol.42 (1995)

## 8.7.2 Semiconductors and Electronic Devices

(See also Sec.3.3.9)

From experience with existing high-energy particle accelerators, it is known that components based on semiconductors are very sensitive to radiation. For example all electronics has been removed from the tunnel of the CERN 450 GeV Super Proton Synchrotron after 6 years of operation at a dose of about 100 Gy. Semiconductors in future multi-TeV collider detectors will have to operate at doses from charged particles exceeding this by orders of magnitude. Therefore the evaluation of electronic components for future detectors requires a well defined radiation test procedure. Specifications for military and space applications [1, 2, 3, 4] can be used, but must be adapted to the radiation environment of high-energy particle accelerators [5, 6, 7].

The degradation in semiconductor devices is caused by ionization and displacement effects. Ionization processes affect the electronic structure of the atoms, causing electron-hole pair generation and resulting in charge accumulation within the insulator and at the interface with the silicon. Displacement damage reduces the charge efficiency and increases the leakage current and noise in the bulk silicon. The degradation can be described by the reduction of the minority carrier lifetime, carrier removal and reduction of the carrier mobility.

The prediction of the radiation hardness of electronic components due to total ionizing dose effects and displacement damage is a very complex task. Since many years, radiation tests have been performed on a large number of various device types from different manufacturers. There are several dependences regarding the radiation hardness of an individual device, such as layout, circuit design, bias condition during and after irradiation or dose rate effects. That means that the test conditions are not always the same. Under this consideration total dose failure levels of relevant technologies and device families are summarized in Fig.1. This review of the sensitivity of the various technologies and device families shows the range at which the *first* failure has been observed (left margin of the bars) to the highest level at which *all* devices failed (right margin of the bars). Fig.1 contains irradiation test results from various test facilities published in recent years. A careful selection by a lot acceptance test is always

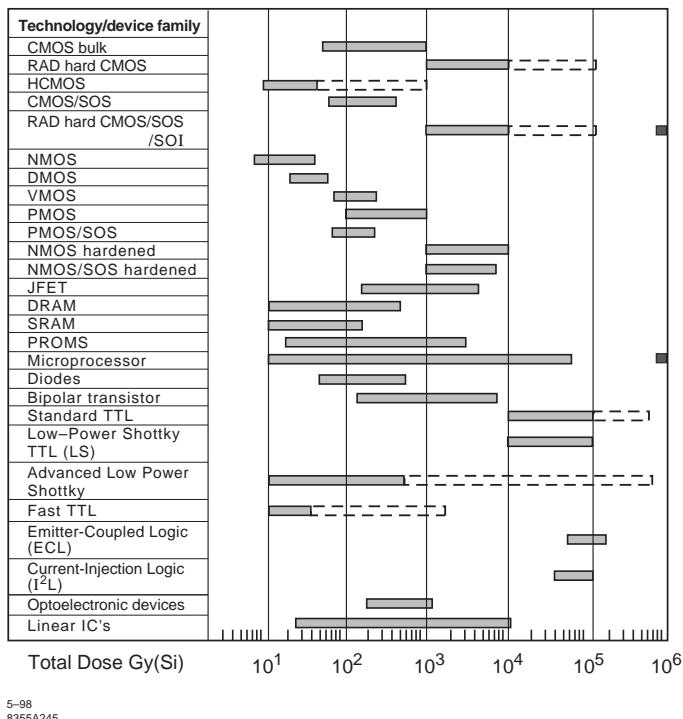


Figure 1: Total dose failure of various technologies and device families. The left bar end indicates the first observed failure and the right hand end the highest level at which all devices failed. Black squares show “ultrahard” technology.

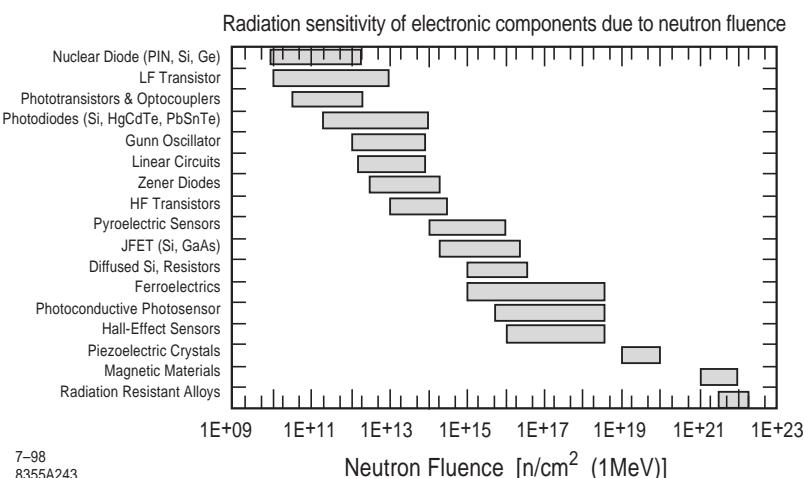


Figure 2: Radiation sensitivity of various components to neutrons.

recommended when the radiation level exceeds the left margin of the bars.

An overview of radiation sensitivity of different components due to neutron fluence is shown in Fig.2. An exhaustive review with a large number of specific data is given in [5].

## References

- [1] A. Boden, F. Wulf and D. Bräunig, Proc. ESA Electronic Components Conference, ESA Publication, Division ESTEC (1990) p.413
- [2] F. Wulf, A. Boden and D. Bräunig, GfW Handbook for Data Compilation of irradiation tested electronic components, HMI Report B-353, TN 53/08, Vol.1-6
- [3] D. Bräunig et al, HMI-B 390, TN 53-10 (1982)
- [4] R.L. Pease, A.H. Johnston, J.L. Azarewicz, IEEE 1990 Int. Nuclear and Space Radiation Effects Conf. (1990)
- [5] F. Wulf et al, Proc. Supercolliders and Superdetectors, World Scientific (1993) p.337
- [6] L. Larsen et al, CERN 93-04 (1993)
- [7] J.E. Gover, T.A. Fischer, IEEE Trans. on Nucl. Sci. 35 (1988) 160

### 8.7.3 Ceramics

Exposure of ceramic materials to ionizing radiation affects the crystal lattice of the ceramic by displacement of atoms within the lattice, conversion of atoms to different species by nuclear reactions, and ionization and electronic excitation.

A variety of effects are produced in ceramic materials as a result of lattice displacement. The introduction of interstitial atoms into the lattice produces swelling which may proceed to the point of causing the material to disintegrate into powder. Defects in the lattice reduce the thermal conductivity of the ceramic. Thermal spikes produced particularly near the end of the path of the particle through the lattice lead to localized regions of high temperatures.

The conversion of atoms in the lattice to other species by reactions with neutrons produces impurity atoms. This is generally not of great importance, except in the cases where gases are formed, such as the neutron reaction with beryllium which produces helium. Such gases in the lattice may nucleate to form gas pockets or bubbles and cause severe swelling. All these effects occur in frequently used structural or insulating materials, such as aluminum oxide ( $\text{Al}_2\text{O}_3$ ) or

beryllium oxide ( $\text{BeO}_2$ ) at very high doses exceeding  $10^{10}$  Gy and neutron fluences above  $10^{21}$  n/cm<sup>2</sup> ( $E > 0.1$  MeV).

Ionization and electronic excitation produced by the passage of fast particles or gamma radiation through the lattice do not affect the general engineering properties of ceramics to any great extent. However, electrons stripped from atoms in the lattice may be trapped at defects to form color centers, or regions of altered optical absorption characteristics. This effect can be of great importance in applications where optical glasses and optical glass fibers are utilized. See next section.

### 8.7.4 Vitreous Materials

Vitreous materials applied in accelerator engineering may be organic or inorganic and are used for windows, optical devices (TV lenses), scintillators or optical fiber cables.

**Glass** Color changes are produced in normal glass during exposure to ionizing radiation. This coloration starts around  $5 \times 10^3$  Gy and is very dark at  $10^5$  Gy.

Types of radiation-resistant glass have been developed which overcome the coloring effects produced by radiation. Glass can be protected from coloring by the addition of certain polyvalent ions called "protective agents" of which cerium is the best known. Addition of 1 to 2%  $\text{CeO}_2$  suppresses the radiation-induced color in cerium glass due to the  $\text{Ce}^{4+}$  ion, which is such a powerful electron acceptor that it removes the radiation-created free electrons in the matter which would otherwise form color centers.

In this way it is possible to produce glasses having no coloration up to  $10^6$  Gy, reduction of transmission between 350 and 550 nm of 10% at  $10^7$  Gy and less than 20% at  $5 \times 10^7$  Gy [1, 2].

**Scintillators** In the last few years a lot of data have been obtained concerning the radiation hardness of organic and inorganic detector scintillating materials for future accelerators. The main conclusions are summarized below [3]:

Radiation-induced damage strongly depends upon the chemical nature of the scintillator's components, the environmental atmosphere and the impurity content.

Among plastics, the matrices Polystyrene (PS) and Polyvinyltoluene (PVT) have shown

the best overall properties. Polymethylmethacrylate (PMMA) is too radiosensitive and must therefore be avoided for applications within ionizing-radiation fields. Polymethylphenylsiloxane (PMPS) needs important improvements in its properties before it can be used as a matrix.

Inorganic scintillators can be arranged in the following order of radiation resistance:  $\text{BaF}_2 > \text{CeF}_3 > \text{CsI}$  (pure)  $> \text{BGO} \gg \text{CsI(Tl)}$  or  $\text{CsF}$ ; the most radiation-hard crystal is  $\text{BaF}_2$ , withstanding dose levels of about  $10^6$  Gy;  $\text{CeF}_3$  and  $\text{CsI}$  (pure) can cope with radiation doses of  $10^4$  Gy; a doped BGO crystal shows improved radiation resistance;  $\text{CsI(Tl)}$  as well as  $\text{CsF}$  are very radiation-soft, and thus of little interest for high-dose level environments.

The decrease in transmission (as the main effect of radiation damage in crystals) has its origin in the formation of color centers. This process is intimately linked with the presence of impurities embedded in the crystalline lattice.

The most radiation-resistant fluors - used for plastic or fluid formulations - are those which emit at the longest wavelengths (e.g. 3-HF), thus bypassing the most affected spectral area of the matrix, or which have a large Stokes shift.

Dopants deliberately added in crystals can preserve (e.g.  $\text{La}^{3+}$  in  $\text{BaF}_2$ ) or increase (e.g.  $\text{Eu}^{3+}$  in BGO) their radiation hardness.

Generally, the presence of air (oxygen) during the irradiation of plastic scintillators was found to have undesirable effects. In the case of liquid scintillators, the presence of oxygen seems to be without consequences.

Recovery exists for PS- and PVT-based scintillators; it strongly depends on the environmental atmosphere and has a beneficial effect on light transmission. In the low-dose-rate range ( $\sim 1$  Gy/h) an equilibrium can be reached between damage and recovery. The latter phenomenon is closely dependent on the size of the plastic scintillator. Recovery also manifests itself in crystalline scintillators. Liquid scintillators show no recovery or, if the phenomenon exists, it is too fast to be evidenced.

Up to now plastic scintillators are able to cope with doses in the  $10^4\text{-}10^5$  Gy range, depending on specific conditions (dose rate, atmosphere, etc.). Liquid scintillators can theoretically withstand higher radiation levels than plastic ones.

Radiation effects in plastic, inorganic and liquid scintillators are extensively discussed in [3].

**Optical fibers** When planning the application of optical fibers in radiation areas of high-energy particle accelerators, a careful assessment of the damage to the fibers caused by radiation, and consequently the selection of the most suitable fiber material for a particular application, are essential. It must be underlined that there is not one best fiber for all purposes. Many boundary parameters have to be taken into consideration.

When optical fibers are exposed to ionizing radiation, they are subject to darkening and hence to an increase of optical attenuation.

Optical attenuation is a result of the absorption of photons by radiation induced color centers. Each particular color center absorbs photons of particular energies, which usually leads to spectral transmission or spectral attenuation.

However, the radiation induced transmission loss depends on many other parameters, such as fiber composition, nature and method of irradiation, time, ambient conditions and fiber "history".

In general, radiation-induced darkening of the fibers is a function of dose and dose rate. Also temperature is an important parameter: induced loss is a decreasing function of temperature.

The transmission wavelength plays also an important role: radiation damage around  $1.3 \mu\text{m}$  is mostly considerably lower than around  $850 \text{ nm}$ .

The recovery is the decrease in attenuation after irradiation, it is also called annealing or fading. If the dose rate is not too high, recovery may take place during irradiation.

The annealing is due to the rearrangement of the molecular or atomic structure of the glass or plastic after darkening. The rate of recovery depends on many parameters, such as: fiber type and composition; fiber drawing process; dose and dose rate of irradiation: high doses lead to fast recovery; fiber "history" (previous irradiations); ambient temperature; if it increases, fading is accelerated; light wavelength; some color centers are more stable than others; light-injected power.

Photobleaching is, as the name indicates, the bleaching of irradiated fiber by injected light. This phenomenon is extensively studied and it seems that bleaching of fiber is accelerated with higher light power. It also depends on light wavelength. It is in many cases recognized that even modest optical power levels can be used to substantially increase the radiation hardness of the fibers.

The luminescence is the temporary light emission phenomenon that takes place in the fiber when it is subjected to radiation; it induces transmission data errors. The luminescence is principally the result of the Cerenkov effect. The intensity of luminescence is a function of fiber composition, of dose rate, of type of irradiation and energy.

The parameters which influence the radiation response of optical fibers can be summarized as follows: type of fiber; dopants (germanium, boron, fluorine, phosphorous); additives and impurities; drawing-induced conditions; irradiation dose; dose-rate; type and energy; temperature; light wavelength; power; in service conditions.

In conclusion, it must be stressed that with such a number of interfacing parameters it is practically impossible to predict the radiation resistance of one fiber or another without proper testing under service conditions.

## References

- [1] R.L. Clough, Radiation resistant polymers, in Encyclopedia of polymer Science and Engineering, Wiley (1988), 20 (13 2nd ed., 1988, p.667)
- [2] B. McGrath, H. Schönbacher, M. Van de Voorde, NIM 135 (1976) 93
- [3] D. Ilie, S. Ilie, H. Schönbacher, CERN-TIS-CFM/IR/91-18 (1991)

## 8.7.5 Metals

The mechanism of radiation effects in metals is essentially related to neutrons which in a primary collision with an atom create vacancies. The neutrons continue to bounce about knocking atoms out of place until their energy is dissipated. A secondary effect results from the energy imparted to the knocked-on atom. This atom collides with other atoms in the lattice knocking them out of position and imparting energy to them. Finally, these "knocked-on" atoms lose all energy and come to rest in an interstitial position in the lattice. Hence, two types of point defects are formed by the collision between neutrons and elastic atoms: vacancies and interstitials.

A further secondary effect of neutron bombardment is the change in structure caused by "thermal spikes". It is postulated that in some instances the energy of the knocked-on atoms can be dissipated locally as a thermal pulse, which can produce extremely high temperatures for very

short periods of time. The area affected by such a "spike" could contain interstitial vacancies, mis-oriented regions and dislocation loops.

Fission processes produce a variety of foreign atoms, the effect of these atoms on interstitial positions will also contribute to property changes. In certain materials, the same effect can be achieved by the transmutation of the basic material to atoms of another element. Of all transmutation products the inert gases cause some of the most notable effects in metals by swelling which is produced by retained gas.

As in the case of ceramics, these effects occur only at very high doses exceeding  $10^{10}$  Gy and  $10^{21}$  n/cm<sup>2</sup> ( $E > 0.1$  MeV).

## 8.7.6 Summary

The major damage mechanisms for various material classes can be summarized as follows:

Semi-conductors: damage due to photoelectric interaction, electron hole pairs and atomic displacements, - damage caused by total dose and by neutrons.

Polymers: damage due to main and side chain rupture, cross-linking degradation, gas evolution, radical production and reaction with the environment - damage caused by total dose.

Ceramics: damage due to displacements, trapped charge carriers, thermal peaks and color centers - damage caused by neutrons and total dose.

Metals: damage due to displacements, dislocations, nuclear reactions producing clusters (He), voids and bubbles. - damage caused by neutrons.

Fig.1 (top, this page) shows for the same materials the radiation dose ranges where they undergo no damage, mild to moderate damage, and where they are usually unusable.

It needs to be noted that: all semiconductor devices with mild to moderate damage shown in dose ranges above  $10^3$  Gy and  $10^{13}$  n/cm<sup>2</sup> are

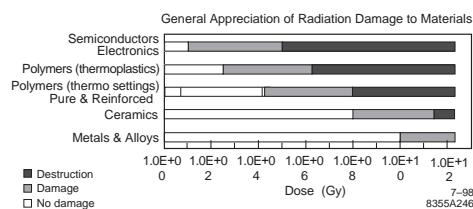


Figure 1: Survey of radiation damage to materials.

## Sec.8.7: RADIATION DAMAGE THRESHOLDS

non-standard products and need to be specially selected for radiation hard technologies; most of the thermoplastic polymers are very sensitive to oxygen-induced damage during irradiations at low dose rates; the limit for use of specially selected organic materials is  $10^8$  Gy; no problems arise in the use of ceramics and metals as structural materials in accelerators.

# Author Index

- Abs, M. 55  
Akai, K. 617  
Aleksandrov, S. 36, 140  
Alessi, J. 568  
Alexander, J. 721  
Alonso, J. 40  
Assmann, R. 300  
  
Baglin, V. 278, 283  
Bai, M. 643  
Balbekov, V.I. 269  
Bane, K. 252  
Bangerter, R. 24  
Barber, D.P. 191, 194, 195  
Bazarov, I. 553  
Belomestnykh, S. 659, 683  
Berg, S. 752  
Berz, M. 105  
Blaskiewicz, M. 199, 396  
Blosser, H. 14  
Bogacz, S.A. 329  
Boine-Frankenheim, O. 739  
Bordry, F. 498, 502  
Boussard, D. 129, 645  
Bowden, G. 451  
Brennan, J. 655, 659  
Brown, K. 71  
Bryant, P.J. 385  
Burkhart, C. 602  
Burnet, J.P. 498, 502  
Burov, A. 85  
Byrd, J. 723, 750  
  
Carrigan, R. 636  
Caspers, F. 732, 745  
Cerdeira Bastos, M. 498, 502  
Chao, A.W. 4, 5, 266  
Chen, P. 175  
Chin, Y.-H. 263  
Chojnacki, E. 473  
Chudakov, E. 756  
Clegg, T. 561  
  
Cleland, M. 22  
Craddock, M.K. 14, 18  
  
Danilov, V. 576  
Decker, G. 709  
Delahaye, J.P. 60  
Douglas, D.R. 329  
Dowell, D. 544, 548  
Dragt, A. 75, 97, 99  
Dugan, G. 565, 621  
Dunham, B. 552  
  
Ecklund, S. 555  
Edwards, D.A. 65-67, 93, 94  
Elleaume, P. 318  
Ellison, J.A. 68  
Emma, P. 334  
Esarey, E. 690  
Evtushenko, P. 719  
  
Faltens, A. 620  
Fang, S.X. 58  
Farkas, Z.D. 481  
Ferrario, M. 133  
Ferry, J. 17  
Fischer, W. 56, 118  
Forck, P. 697, 699, 702, 704  
Fox, J. 628, 649  
Furman, M.A. 115, 163, 311, 576  
  
Gai, W. 754  
Gamp, A. 653  
Ganni, R. 425  
Gareyte, J. 752  
Garoby, R. 376  
Gaskell, D. 756  
Gjonaj, E. 248  
Gluckstern, R.L. 243  
Goddard, B. 511  
Gollwitzer, K. 8  
Grames, J. 756  
Gröbner, O. 278  
Gundersen, M. 533

- Haeberli, W. 561  
Halbach, K. 584-586, 588, 607  
Hammer, D.A. 51  
Hangst, J.S. 205  
Harding, D. 577  
Hartill, D. 12, 40  
Hays, T. 420  
Hellborg, R. 27  
Henderson, S.D. 36, 140, 286, 516  
Hinkson, J. 706, 709  
Hirata, K. 169  
Hoffstaetter, G. 127  
Huang, H. 642  
Huang, Z. 229
- Irwin, J. 367, 370  
Ispiryan, K. 238
- Jain, A.K. 523  
Jing, C. 754  
Jones, R. 362, 736  
Jongen, Y. 55  
Jowett, J.M. 180
- Keil, E. 77  
Keizer, R. 639  
Kelly, M. 681  
Kim, K.-J. 318  
Kirchgessner, J. 683  
Kirk, H. 543, 547  
Kornilov, V. 739  
Koutchouk, J.-P. 391  
Krafft, G.A. 235, 329  
Krinsky, S. 227, 350  
Krisch, A.D. 187  
Kubo, K. 123  
Kuchnir, M. 399  
Kurennoy, S.S. 243, 265
- Lebedev, V. 85, 155  
Lengeler, H. 47  
Leonova, M.A. 187  
Leung, K. 563  
Li, Y. 431-433, 435, 439, 440, 445  
Loew, G.A. 34, 507, 664  
Löhl, F. 763
- Mahner, E. 278  
Mais, H. 68  
Makdisi, Y. 758  
Markley, F. 410  
Marks, S. 529  
Marriner, J.P. 8  
Maruyama, T. 546
- McAshan, M. 406, 414  
McMahan, M. 571  
Mikhailichenko, A. 558  
Mills, F. 577  
Mistry, N.B. 431-433, 435, 439, 440, 445  
Mokhov, N.V. 269, 293  
Montag, C. 453, 459
- Nagaitsev, S. 202, 396  
Nantista, C. 486  
Nation, J. 51  
Neuffer, D. 209  
Ng, K.Y. 116, 152, 252  
Noda, A. 205
- Oide, K. 277
- Padamsee, H. 420, 674  
Palmer, R. 44  
Paret, S. 739  
Peterson, D. 721  
Phillips, H.L. 447, 491, 515  
Pikin, A. 568  
Piwinski, A. 94  
Plum, M. 574  
Poelker, M. 546  
Potter, J. 660  
Proch, D. 420, 494, 674  
Ptitsyn, V. 641, 643
- Raubenheimer, T.O. 321, 329  
Rees, G. 382  
Rehlich, K. 760  
Rice, D. 346, 347  
Rimmer, R.A. 521, 672  
Ripken, G. 68, 191, 194, 195  
Rode, C. 425  
Roesler, S. 767, 769, 776, 779, 786, 790  
Rosenzweig, J. 31  
Roser, T. 183-186, 643  
Ross, M. 726  
Rossbach, J. 453  
Roth, G. 533  
Rubin, D. 352, 365  
Ruland, R. 448  
Ryne, R. 62
- Sabbi, G. 599  
Safranek, J. 357  
Sanders, J.M. 533  
Scanlan, R.M. 404, 465  
Schlueter, R.D. 471, 529, 584-586, 588, 607  
Schmidt, B. 713  
Schmüser, P. 229, 588

- Schnell, W. 670  
Schönbacher, H. 793  
Schreiber, S. 326  
Schroeder, C. 690  
Schulte, D. 175  
Seeman, J.T. 381, 716  
Sekutowicz, J. 494, 550  
Seryi, A.A. 337  
Shaposhnikova, E. 742  
Shiltsev, V. 391, 394, 641  
Shintake, S. 728, 730  
Silari, M. 767, 769, 776, 779, 786, 790  
Smith, S. 709  
Spencer, J.E. 273, 275  
Sprehn, D. 651  
Staples, J. 52  
Steier, C. 81  
Steinhagen, R. 362  
Strehl, P. 697, 699, 702, 704  
Striganov, S.I. 293  
Stupakov, G.V. 121, 167  
Sullivan, M. 286  
Sundelin, R.M. 491, 515  
Suzuki, T. 242, 262  
Symon, K. 89  
Syphers, M. 65-67, 93, 94, 360  
  
Tang, T. 533  
Tantawi, S. 485, 486, 507, 540  
Thomas, R. 769  
Thompson, K. 123  
Thongbai, C. 226  
  
Tian, Y. 624  
Tigner, M. 4, 10, 266, 411, 412, 414, 463, 471, 475, 496, 521, 617, 653  
  
Venturini, M. 75  
  
Wanderer, P. 523  
Wang, J.W. 507  
Wei, J. 47, 212, 373  
Weiland, T. 248  
Welch, J. 614  
Wenninger, J. 350  
White, G.R. 337  
Whittum, D. 687  
Wiedemann, H. 215-217, 220, 223, 237, 588  
Wilson, P. 263  
Wittenburg, K. 736  
Woods, M. 756  
Wolski, A. 111  
  
Yan, Y.T. 367, 370, 373  
Yokoya, K. 123, 198  
Yoshida, H. 109  
Yoshii, Y. 667  
Yu, L.H. 624  
  
Zeller, A. 595  
Zhang, C. 58  
Zholents, A. 233  
Zimmermann, F. 159, 286, 387, 732  
Zisman, M.S. 271, 311  
Zobov, M. 174, 387  
Zotter, B. 137, 144, 145, 147, 150, 263

**This page intentionally left blank**

# Subject Index

- Aberrations 83, 98, 102, 337-346 (see Transport matrices, Maps, Multipoles)  
plasma lens 32  
electrostatic lens 620  
lithium lens 623
- Abbreviations 1
- ABS (acrylic plastic) 308, 464, 795
- Absorbers (rf) (see Lossy materials)
- Absorbed dose 767, 768, 771
- Accelerating structures  
constant gradient, constant Impedance 35, 665, 666  
coupled-cavity 35, 38  
detuned 34, 126  
drift tube (see Drift tube linac)  
elliptical superconducting 38  
ferrite loaded (see Inductively loaded accelerating cavities)  
millimeter wave 687-690  
multicell normal conducting 670-672  
multicell superconducting 674-681  
RFQ (see Radio frequency quadrupole)  
scaling with frequency 665  
single cell normal conducting 672-674  
single cell superconducting 683-687  
spoke cavity 38  
standing wave 34, 54, 123, 478, 618, 664-667, 670-672, 674-681  
traveling wave 34, 123, 618, 664-667  
 $v_p < c$  linac 660-664, 681-683  
 $v_p = c$  linac 34-36, 664-667, 674-681
- Acceleration subsystems 645-697
- Accelerator applications (see Industrial accelerators, Medical accelerators)
- Accelerator driven subcritical (ADS) reactor 19, 50
- Activation detectors 792
- Activity 767
- Adapter (see Emittance exchange)
- Adiabatic damping (see Damping, adiabatic)
- Adiabatic invariant (see Invariant)
- Adiabaticity (see Rf gymnastics)
- Admittance 65, 300-308, 382-387, 558
- Air (properties) 308, 417, 432
- Alfvén current 133
- Alignment 448-451  
beam-based 324, 343, 346-349, 357-360 (see Supports)
- Alpha magnet 588
- Alternating gradient focusing (see Focusing)
- Aluminum 308, 310, 400-403 (see Vacuum chamber)
- Aluminum nitride 463, 465
- Aluminum Oxide, Alumina 400, 447, 448, 463, 465
- Ambient dose equivalent 768, 779-790
- Amplitude function (see Courant-Snyder functions)
- Anomalous magnetic moment 184
- Anomalous skin effect (see Skin depth)
- Antihydrogen 10
- Antiproton source (see Sources)
- Antiproton production cross-section 8, 565 (see Sources, antiproton)
- Arnold's diffusion (see Dynamic aperture)
- $A * T * L$  relation 324, 454
- Atomic and nuclear properties, table 308, 310
- Atomic beam source 561
- Autophasing (see Collective Instabilities)
- Backgrounds 178, 286-300  
 $e^\pm$  collider 178, 286-290, 293-298  
muon collider 298  
shielding 288, 290, 296-298  
showers 293-296  
simulations (see Computer codes)
- Baker-Campbell-Hausdorff (BCH) theorem (see Lie algebra)
- Band structure, multicell cavity 678
- Barrier bucket (see Rf gymnastics)
- Bayard-Alpert gauge (see Vacuum, gauge)
- Bead pull 54 (see Cavity, measurements)
- Beam based alignment (see Alignment)

- Beam-beam compensation 391-394  
electron lens 391, 641  
four-beam compensation 391  
octupole compensation 392  
phase advance between IPs 392  
alternating crossing 392  
long-range wire compensation 392  
long-range multipole compensation 393
- Beam-beam effects, linear collider 175-180  
background (see Backgrounds)  
beamstrahlung (see Beamstrahlung)  
depolarization (see Polarization)  
disruption 176, 178  
disruption parameter 176  
hadron production 179  
kink instability 176  
luminosity enhancement (see Luminosity)  
multi-bunch instability 176, 339  
pair production (see Pair production)
- Beam-beam effects, storage ring colliders 169-174, 180-183  
Bassetti-Erskine formula 169, 181 (see Space charge)  
beam separation criteria (see Parasitic beam-beam effects)  
beamstrahlung (see Beamstrahlung)  
coherent beam-beam effects 171, 172  
crab crossing (see Crab crossing)  
crab waist 174, 175, 313, 387-391  
dynamic beta 170  
dynamic emittance 170  
elastic scattering 362  
hour-glass effects (see Hour glass effect)  
Large Piwinski Angle scheme 389  
long range (see Parasitic beam-beam effects)  
long-range beam-beam effects (see Parasitic beam-beam effects)  
Pacman effect 183  
parasitic collision (see Parasitic beam-beam effects)  
Piwinski angle 80, 174  
round beam collisions 88, 389  
synchrobetatron resonances, crossing angle 96, 173  
transparency conditions 311  
traveling focus 389  
tune shift, parameter 96, 169, 182  
tune shift with amplitudes 171  
with crossing angle 96, 173, 174  
with dispersion at IP 173, 174
- Beam-beam elastic scattering (see Emittance, beam-beam)
- Beam break-up (BBU) 34, 123-127, 323 (see Collective instabilities)  
beam loading (see Beam loading)  
BNS damping (see Collective instabilities)  
control multibunch BBU 126  
cumulative 125  
daisy chain 126  
emittance dilution due to rf structure mis-alignments 127  
in ERLs 127-129  
quadrupole mode 125  
recirculative 128  
two-particle model (see Collective instabilities)
- Beam bunching (see Rf gymnastics)  
velocity (see Velocity bunching)
- Beam current (see Current)
- Beam dump, linear colliders 345
- Beam ellipse 72
- Beam-gas scattering (see Backgrounds, Beam lifetime, Emittance, beam-gas, Scattering cross-section, and individual scattering cross-sections)
- Beam halo 48, 143, 298, 323, 329
- Beam-ion instability (see Ion trapping)
- Beam lifetime  
against thermal photons (see Thermal photons)  
bremsstrahlung with nuclei and dust 293  
electrons 271-273  
protons 269-271  
residual gas 269-271
- Beam loading 123, 129-133, 327, 619, 665-670  
AM/FM modulation transmission 130  
pbar collection 567  
compensation, linacs 123, 322, 327  
compensation, storage rings 132  
equivalent circuit 130  
fundamental theorem 129, 242  
minimum energy spread, linacs 124, 667  
plasma accelerators 691-697  
transient 123, 132, 672
- Beam position centering  
quadrupoles 348  
sextupoles 348
- Beam position measurement (see Position measurement)
- Beam position monitor, BPM
- Beam profile measurement (see Size measurement)
- Beam size measurement (see Size measurement)
- Beam spectra (see frequency spectra)

- Beam transfer functions, BTF 396, 739, 742
  - longitudinal bunched 742
  - longitudinal unbunched 742
  - transverse bunched 740
  - transverse unbunched 740
- Beam tubes (see Vacuum systems)
- Beamstrahlung 44, 177, 178 (see Beam-beam effects, linear colliders), 289
  - average energy loss 177
  - number of photons 177
  - parameter 177
- Becquerel 767
- Bent crystal (see Crystal bending)
- Beryllium 308, 310, 400-403, 414
  - pipe 445
- Beryllium oxide, Beryllia 442, 463, 465
- Beryllium-copper 400, 442, 445
- Bessel functions, properties 5
  - roots 260, 266, 478, 670
- Beta function (see Courant-Snyder functions)
- Betatron 10-12
  - 2-to-1 condition 11
  - separated-function 11
  - modified 11
  - FFAG 12
- Betatron function (see Courant-Snyder functions)
- Betatron motion 65, 66
  - equation of motion 65
  - transport matrices (see Transport matrices, Maps)
- Betatron phase 65
  - dependence on beam current 752
- Betatron tune (see Tune)
- Bethe-Bloch equation 209, 267
- Bethe-Heitler equation (see Bremsstrahlung)
- Bethe small hole impedance (see Impedance, beam)
- BF<sub>3</sub> counters 786, 787
- BGO 308, 789
- BGO photon spectrometer 789, 800
- Bhabha scattering 272, 274
- $B_{\text{mag}}$  factor 322
- Black body radiation photons (see Thermal photons)
- Blow-out regime 692
- BMT equation (see Thomas-BMT equation)
- BNS damping (see Collective instabilities)
- Boiling point 308, 310, 407, 408
- Bonner sphere 787
- Bonner sphere spectrometers 787
- Boron neutron capture therapy (BNCT) 20, 43
- Boussard criterion (see Landau damping)
- Brachytherapy 43
- Bragg peak 41
- Brazing 443, 447, 448
- Breakdown
  - high voltage 51, 396, 511-515, 533, 536, 552, 574, 603, 615, 620
  - rf 31, 34, 44, 507-511, 508, 509, 548, 653, 663 (see Kilpatrick criterion), 688
  - thermal 420
- Bremsstrahlung 238, 287, 341, 555, 769, 770
  - attenuation 616, 781
  - beam-beam 268 (see Bhabha scattering, radiative, Beam-beam effects, linear collider)
  - coherent 239
  - dust particle 161
  - gas nuclei 161, 268, 272
  - muons 293
- Brightness 81-85, 216, 318-321
  - particle beam 318
  - radiation beam 319
  - Wigner distribution 319
- Brillouin diagram 34
- Brillouin flow 135
- Bubble detectors 787
- Bucket (see Rf bucket)
- Buckling 414
- Bunch compression 66, 334
  - CSR effect 336
- Bunch distribution functions (see Size measurements, Potential well distortion) 124, 227, 263, 617, 699, 702, 713-721, 723, 726, 728, 730
- Bunch length, radiative equilibrium 221
- Bunch lengthening (see Collective instabilities, microwave instabilities)
  - Haissinski equation (see Haissinski equation)
  - potential-well (see Potential well distortion)
  - scaling law 146
  - turbulence 146, 148, 152
- Bunch manipulation (see Rf gymnastics)
- Bunchers 34-36, 44, 53, 548, 554, 700
- Buncher voltage 554
- Busch's theorem 135
- Cam mover 452
- Cancer induction 769
- Capture condition (see Electron preinjector)
  - longitudinal phase trajectory, e<sup>+</sup> collection 557
- Cascade, electromagnetic (see Showers)
- Cascade generators (see Electrodynamic accelerators)

- Cathodes (see Electron sources)
  - field-induced (see Field emission)
  - thermionic 543, 544
  - photo 544-547
  - photo, longevity 554
  - dispenser 543, 544
- Causality (see Wakefields)
- Cavity (see Accelerating structures)
  - modes 147, 478-481
  - coupling (see Couplers)
  - measurements 521, 522
- Ceramic vacuum chambers 447, 448, 516
  - metalization 447
- Channeling 636-639
  - Lindhard angle 240, 636
  - radiation 240
  - crystal extraction 384
- Chaos (see Dynamic aperture)
- Characteristic impedance, transmission line (see Impedance, transmission line)
- Charge exchange injection (see Injection)
- Charge state determination 704-706
- Charge state strippers (see Strippers)
- Chasman-Green lattice (see Lattice)
- Cherenkov radiation 238
  - coherent 32
- Child's law (see Electron sources)
- Chirikov criterion (see Dynamic aperture)
- Chlorofluorocarbon gases 409
- Chromaticity
  - correction, final focus 339
  - correction, storage rings 83 (see Dynamic aperture)
  - FODO cell 78, 229
  - natural 66
  - nonlinear 93, 103
  - real time measurement & control 362-365
- Closed orbit (see Orbit distortion, Orbit correction)
- Coalescing (see Rf gymnastics)
- Coatings
  - rf windows 515, 516
  - metal and ceramic chambers 516-521
- Coaxial lines (see TEM lines)
- Coaxial waveguide (see Waveguide)
- Cockcroft-Walton accelerator 23, 36, 512
- Codes (see Computer codes)
- Coercive force 472, 579
- Coherence
  - spatial coherence 82, 217
  - temporal coherence 217
- Coherent bremsstrahlung radiation (see Bremsstrahlung)
- Coherent synchrotron radiation 167, 168, 216, 227 (see Synchrotron radiation)
  - as measurement tool 713
  - CSR instability 118, 168, 228
  - curvature effect 168
  - formation length 168
  - potential-well distortion 228
  - wall shielding 168
- Coherent electron cooling (see Stochastic cooling)
- Coils (see Magnet)
- Cold cathode gauge (see Vacuum, gauge)
- Collapse (see Buckling)
- Collars, magnet 590
- Collective effects 123-168
- Collective instabilities 123-168 (see Impedance, beam, Wakefield)
  - beam break-up (see Beam break-up)
  - BNS damping (autophasing) 124, 322
  - Boussard criterion (see Landau damping)
  - coherent synchrotron radiation (see Coherent synchrotron radiation)
  - coupled bunch, CBI 150-152, 725, 750-752
  - dispersion relation (see Landau damping)
  - energy recovery linac 127-129
  - fast wave, slow wave 354
  - head-tail 148, 154, 725, 732, 740
  - Keil-Schnell criterion (see Landau damping)
  - Landau damping (see Landau damping)
  - linac, electron 123-127
  - linac, proton 140-144
  - longitudinal head-tail 117, 149, 155
  - longitudinal mode-coupling 151, 155
  - microwave instabilities 148, 152, 374
  - negative mass 137, 147
  - resistive wall 148, 149, 151
  - Robinson instability 147, 646
  - strong/fast head-tail 149
  - transverse mode-coupling (TMCI) 149, 154
  - turbulence (see Bunch lengthening)
  - two-particle models 124, 149
- Colliders 12-14
  - linear 12, 36, 298, 321-329, 337-346
  - muon (see Muon collider)
  - storage ring 12, 77, 346-362
- Collimation 300-308
  - crystal 638
  - magnetized collimator 298
  - tip scattering 288
  - wakes 252, see Impedance, collimator

- Collision schemes 387-391 (also see Beam-beam effects)  
Columbium (see Niobium)  
Combined function lattice (see Lattice)  
Compaction factor (see Momentum compaction factor)  
Composition diagnosis, ion beam 697-699  
Compressor (refrigeration) 426, 427  
Compton scattering 275-277, 288, 757  
    against thermal photons (see Thermal photons)  
    back scattering 235  
    inverse 31  
    Klein-Nishina formula (see Klein-Nishina cross-section)  
    multiple Compton scattering 276  
Computer codes 62-64  
    beam dynamics 63  
    beam-beam simulation 175, 182  
    beam-material interaction 64  
    community standard 62  
    coupled optics 87  
    detector background simulation 274, 291  
    differential algebra 107  
    DTL design 38, 662  
    dynamic aperture 112  
    electron gun 543  
    electromagnetics 63  
    final focus tuning 343  
    free electron laser 64  
    HOM cavity structure 63  
    ionization cooling 211  
    impedance 63 (see Impedance, calculation)  
    particle tracking simulation 112  
    particle loss simulation 291  
    polarization, electron beam 194, 195  
    proton linac design 144  
    radiation shielding 776  
    rf cavity 665  
    RFQ design 53, 54  
    synchrotron radiation modeling 64  
Conductance, tubes and orifices 433-435  
Conduction (see Heat transfer)  
Conductivity  
    ac conductivity  
    electrical (see Material properties, electrical)  
    thermal (see Thermal conductivity, Material properties, thermal)  
Confinement (see Magnet, Lenses, Feedback, Separators)  
Constant gradient rf structure (see Rf structures)  
Constant Impedance rf structure (see Rf structures)  
Continuity equation 6  
Control systems 760-763  
    architecture 760  
    availability 763  
    network 761  
    software 762  
Convection (see Heat transfer)  
Conversion of units (see Units)  
Coolant properties 406-410, 416, 417, 616  
Cooling, beam  
    coherent electron (see Stochastic cooling)  
    crystal 638  
    electron (see Electron cooling)  
    ionization 20, 45, 209-212  
    laser 205-209  
    stochastic (see Stochastic cooling)  
Cooling tower 419  
Coordinate system 65, 89  
    Cartesian, cylindrical, spherical 6  
    Frenet-Serret 6, 69  
Copper 308, 400-402, 441, 475, 545, 556 (see Vacuum chamber)  
Cores, magnet (see Magnet and Induction accelerator)  
Corrections (see Error corrections)  
Corrosion 419  
Cosmic ray background radiation 769  
Coulomb force 6  
Coulomb log 156, 157, 203 (see Intrabeam scattering)  
Coulomb scattering 266, 287, 361  
    multiple 267  
Coupling, particle motion  
    coefficient 222  
    coherent motion 87  
    4-D ellipsoid 86  
    Edwards-Teng parametrization 86  
    eigenanalysis 68-71, 85, 353, 354  
    eigen-emittance 86, 105  
    equation of motion 68  
    Mais-Ripken parametrization 86  
    measurement 354, 355, 358, 363, 371  
    perturbative theory 87  
    real time measurement & control 362-365  
    sources 352  
    synchrobetatron 68, 92, 93, 94-97  
    transverse 67, 85-89, 91, 352-357  
    tune split 354  
Coupling, cavity 493

- Coupling correction (see Error sources and corrections, Operational considerations)
- solenoid compensation (see Solenoid)
- Couplers
  - fundamental mode 493, 673-681
  - higher order modes 475, 674-681
- Courant-Snyder functions  $\alpha, \beta, \gamma$  65, 72
  - $\beta$ -function and phase measurements 349, 359
  - $\beta$ -wave 346, 360
  - corrections (see Operational considerations)
  - FODO cell 78
  - near a waist 79
- Courant-Snyder invariant (see Invariant)
- Crab crossing 176, 313, 339, 389, 617
- Crab cavity 617 (see Separators, rf)
- Crab waist (see Beam-beam effects, storage ring colliders)
- Creep 410, 411
- Critical energy, particle-matter (see Electron)
- Critical energy, synchrotron radiation (see Synchrotron radiation)
- Critical field (see Compton scattering)
- Critical point 407
- Critical properties (fluids) 407, 408
- Cross-sections (see Scattering cross-sections)
- Crossing angle (see Beam-beam effects) (see Luminosity)
- Cross transformer 552
- Cryogen
  - properties 406-410
  - refrigerators 425-429
  - storage 429
  - supercritical flow 430
  - transfer lines 429-431
  - two phase flow 430
- Cryostat, cavity 326, 683-687
- Crystal bending (see Channeling)
- Crystalline beams 212-215
  - condition 212
  - experiments 214
  - ground state 212
  - Hamiltonian 212
  - lattice heating 213
  - molecular dynamics method 212
- Curie 4, 767
- Current measurement, beam 706-709
- Current sheet equivalent material (see Magnetic materials)
- Current transformer (see also DCCT) 503, 706-709
- Curvature effect (see Impedance)
- Curvilinear coordinates (see Coordinate system)
- Cyclotron 14-17
  - extraction 16
  - FFAG (see FFAG accelerator)
  - flutter 16, 19
  - $H^-$  15
  - isochronous cyclotron 14
  - resonances (see Resonances)
  - sector-focused 15, 27
  - space charge (see Space charge)
  - superconducting 15
  - synchrocyclotron 15, 18, 27
  - Thomas cyclotron 15
- Cyclotron frequency 14
- Cylinder model of multipoles (see Multipole expansion)
- $\delta$ -rays 293
- Damping (see Cooling)
  - adiabatic 66, 67, 134
  - radiation 70, 220
  - vibration 459-461
  - wigglers 222
- Dark current 36, 508, 639, 675
- DBA Double bend achromat (see Lattice)
- DCCT 503, 706
- Debunching (see Rf gymnastics)
- Dechanneling 637
- Decibel, dB 4
- Decoherence 115, 116, 118 (see Echo)
- Deep inelastic interactions 294
- $dE/dx$  (stopping power) 267, 293, 308, 774
  - values 308, 736
- Deflection (mechanical) 412, 413, 513
- Deflectors (see Separators)
- Density, values 308, 407, 408, 463, 464
- Depolarization (see Resonances, depolarization, Polarization)
- Desorption, beam induced and thermal (see Outgassing)
- Detuned rf structure (see Rf structure)
- Detuning effect, betatron oscillation 93, 119
- Deuterons 3, 15, 188, 562, 775, 792
- Diagnostics, beam 697-745,
- Dielectric constant (see material properties)
- Dielectric laser accelerator (DLA) 31
- Dielectric strength (see material properties)
- Dielectric wakefield accelerator (see Wakefield accelerator)
- Dielectrics (properties) 308, 463-465
  - fluids 616
- Differential algebra DA 105-109
  - n-th order integrator 106

- symplectic tracking 107
- truncated power series algebra (TPSA) 106
- Diffraction limit 81, 216, 232
- Diffraction model (see Impedance, beam)
- Diffraction radiation 227
- Diffusion 118-122, 145, 361, 362, 646 (see Dynamic aperture, Noise)
  - thermal diffusion 453
  - tune 114, 119
- Dipoles (see Magnet, Separators, Optics, Operational considerations)
- Dispersion function 66
  - errors 346, 360, 371 (see Alignment, beam based)
  - FODO cell 78
  - measurement 348, 365
  - nonlinear in  $\delta = \Delta P/P_0$  117
- Dispersion relation (see Brillouin diagram, Landau damping)
- Dispersion suppressor (see Lattice)
- Disruption (see Beam-beam effects, linear colliders)
- Distortion (see Deflections)
  - electrostatic 513
  - thermal 452
- DLDS see RF Distribution
- Doors (see Shielding)
- Doppler effect 216 (see Cooling, laser)
- Dose conversion coefficient 768
- Dose equivalent 768, 769, 775, 779-790
- Dosimetry
  - muons (see Muon, monitoring)
  - neutrons (see Neutron, monitoring)
  - photons (see Gamma, monitoring)
  - protons (see Proton, monitoring)
- Double bend achromat (see Lattice)
- Double slit interferometer (see Laser interference methods)
- Doublet, quadrupoles 79
- Drift tube linac DTL 36-40, 660-664
- Ducts (see Shielding, fluid flow)
- Ducts and Labyrinths 782, 783
- Dust trapping 160, 161
- Dynamic aperture 83, 111-115 (see Nonlinear dynamics)
  - amplitude diffusion 120
  - Arnold's diffusion 112
  - chaos, tune modulation 118, 121
  - Chirikov criterion 112, 121
  - distortion 119
  - factors affecting dynamic aperture 112
  - frequency map analysis 114, 119, 175
- KAM theorem (see KAM theorem)
- long-term predictions 113-115
- Lyapunov exponent 114
- measurement 113 (see Nonlinear dynamics, experiments)
- optimization 114
- ripple and tune modulation effects 112, 120
- smear 118
- survival plot 113, 120
- swamp plot 115
- Dynamic beta (see Beam-beam effects, storage ring colliders)
- Dynamitron 23
- Earth shielding (see Shielding)
- Earth vibrations (see Vibration)
- EBIS 568
- Echo 118, 121-123 (see Decoherence)
  - echo-enabled harmonic generation (EEHG) (see Free electron laser)
  - diffusion effects 122
  - experiments 122
  - longitudinal 122
  - spin 122
  - transverse 121
- ECR 570
- Eddy currents 11, 347, 383, 412, 420, 581, 586, 587
  - core saturation 587
  - decay (rise) time 517, 592
  - field distortions 592
  - flux concentrator 557
  - forces 412
  - heating 517, 579, 581, 592, 596, 668
  - ripple shielding 347
  - septa 641
  - in superconducting cables 467
- Edge radiation 237
- EEHG (see Echo)
- Effective dose 768, 769
- Eigenanalysis
  - emittance (see Coupling, particle motion)
  - linear coupling (see Coupling, particle motion)
  - orbital motion SLIM 194, 195
  - spin motion SLIM (see Polarization, radiative)
- Ejection (see Injection methods)
- Elastance 663
- Elastic modulus, Young's modulus, values 310, 405, 402, 463, 464, 795
- steel 580

- Electrical conductivity (see Material properties, electrical)
- Electrode materials 512, 616
- Electrodynamic accelerators (see High voltage)
- Electromagnetic cascade (see Showers)
- Electron
  - bremsstrahlung (see Bremsstrahlung)
  - critical energy 238, 266, 268
  - neutron production by 770, 772
  - production (see ParticleSources)
  - radiation length (see Radiation length)
- Electron beam welding (see Welding)
- Electron-cloud effect 50, 163-167
  - diagnostics 732-736
  - head-tail instability 165
  - mitigation 165
  - multipacting (see Multipacting)
  - photoelectron instability 163
  - photoemission 283, 284
  - pressure rise 165
- Electron cooling 8, 9, 88, 202-205, 396
  - cooling rates 203
  - friction force 203
  - magnetized 203
  - operational considerations 396
  - relativistic electron cooling 204
  - secondary emission 284, 285
  - two-component model 202
- Electron guns 134 (see Particle Sources, Electron cooling)
  - Child's law 134, 543, 656
  - DC gun 134
  - emittance scaling 137
  - perveance 135, 543, 652
  - Pierce gun 134, 550
  - rf gun (see Rf guns, electron)
  - Richardson's law 543
  - simulations 548
  - rf guns (see Rf guns)
  - space charge effects 133 (see Space charge)
- Electron preinjector 547, 553 (see Particle Sources, electron)
  - buncher voltage 395, 554
  - dynamics, longitudinal 557
  - dynamics, transverse 134 (see Envelope equation)
  - sources of emittance (see Emittance)
- Electron ring accelerator 24
- Electropolishing (see Niobium, electropolishing)
- Electrostatic accelerators 17, 18
- Electrostatic lens 620, 621
  - acceleration tube 17
- charging system 18
- support column 18
- Electrostatic separators (see Separators)
- Emissivity, various materials 418
- Emittance 65, 72
  - beam-beam 362
  - beam-gas 323, 341, 361
  - bumps 325
  - compensation, gun 136
  - correction schemes, linear colliders 325
  - diffraction limited (see Diffraction limit)
  - dilution, linacs 124, 125, 322-324
  - dilution, storage rings 360-362
  - dipole noise (see Noise)
  - dispersion-free steering 325
  - eigenemittance (see Coupling, particle motion)
  - equipartition 144
  - feedback noise 629
  - growth in transition crossing (see Transition)
  - halo (see Beam halo)
  - intrinsic 545
  - longitudinal 67, 319
  - measurement 702-732
  - minimum, DBA, TBA, multi-bend lattices (see Lattice)
  - minimum emittance theorem 105
  - normalized 66, 319
  - oscillation, gun 135
  - radiation equilibrium 70, 82, 145, 221
  - rf bucket mismatch 361
  - rf noise (see Noise)
  - sources, electron gun (see Electron guns)
  - space charge induced, gun 134
  - thermal 134
  - wake-free steering (see Wake free steering)
- Emittance exchange 68, 87
  - applications 88
- Emittance measurement (see Size measurement, beam)
- Emulsions, nuclear 787
- Energy measurement, electrons 193, 381, 382, 643
- Energy mismatch 360
- Energy recovery linac 329-334
- Energy spread (see also individual accelerators)
  - electron linac 321, 322, 667
  - electron storage ring 221
- Energy spread compensation
  - linacs 123, 124, 667
  - monochromatization (see Monochromatization)

- Envelope equation 134, 138, 142
  - emittance dominated, space charge dominated 133
- Kapchinskiy-Vladimirskiy (KV) distribution 142
- Equivalent dose 767
- Error function,  $\text{erf}(x)$ , properties 5 (see Beam-beam effects, Bassetti-Erskine)
- Error sources and corrections (see Operational considerations, Emittance dilution)
  - beta function 346
  - coupling 352-359
  - dispersion 346, 357
  - $\epsilon$ -bump (see Emittance, bumps)
  - error distribution, expected 324
  - orbit 346, 350-352, 357
  - real time measurement & control 362-365
  - tune 346
- Etch track detectors 788
- Evaporative cooling (see Cooling tower)
- Expander (refrigerator) 428
- Explosion bonding 444
- Exposure (radiation) 472, 767-799
- Extraction 385-387 (see Injection, individual devices)
  - cyclotron 16
  - Hardt condition 385
  - $H^-$ , cyclotron 15
  - single-turn extraction 384
  - fast multiturn extraction 384
  - multi-turn crystal extraction 384
  - ripple effects 386
  - slow extraction 42, 118, 384, 385
- Faraday cup 708
- Fast beam-ion instability (see Ion trapping)
- Feedback
  - bunch by bunch (coupled bunch) 628-636
  - global orbit 624 (see Orbit correction)
  - local orbit 624
  - rf 647, 649-651, 680
  - tune, coupling and chromaticity 362-365
- Feedforward rf 647, 680
- Ferrite 471, 472, 474, 607, 667-670
- FFAG accelerator 18-22
  - muon source, acceleration 20
  - nonscaling 20
  - quasi-scaling 20
  - scaling 19
  - vertical scaling 20
- Field emission 36, 507, 513, 689
- Field index 19
- Filters 201, 629, 636, 648, 650, 672
- Final focus, linear collider 337-346
  - machine-detector interface 345
  - operation 344
  - tolerances 341
  - tuning 343
- FINEMET® 668
- Finite integration theory FIT 249
- Fluence 767
- Fluid flow (see Heat transfer)
- Flutter (see Cyclotron)
- Focusing 577
  - alternating gradient, strong focusing 59
  - combined function, separate function (see Lattice)
  - weak focusing 59
- FODO cell (see Lattice)
- Fokker-Planck equation 71, 145
- Foil strippers (see Stripping)
- Forces (electromagnetic) 411, 412, 587, 588, 590
- Fourier equations, heat transfer 415
- Fowler-Nordheim emission (see Field emission)
- Fraunhofer integral 216
- Free electron laser, FEL 229-233, 229
  - EEHG mode 122, 233
  - FEL parameter 231
  - gain length 231, 232, 229
  - HGHG mode 233
  - HHG mode 233
  - microbunching 227, 230
  - Rayleigh range (see Rayleigh range)
  - SASE mode 231
  - self-seeding 233
- Frequency control, proton synchrotron 645
- Frequency map analysis (See Dynamic aperture)
- Frequency spectra (see Beam transfer function, Schottky spectra)
  - multiple bunches 750-752
- Fringe factor (hydraulic) 416
- Fringe field (see Magnet, pole shaping)
  - maps (see Maps)
- Froissart-Stora equation 186, 189
- Fundamental theorem, beam loading (see Beam loading)
- Fundamental constants 3
- G-10 400-403, 795
- GaAs 545, 546
- Gain length (see Free electron laser)
- Gamma function  $\Gamma(x)$ , properties 5
- Gamma radiation
  - monitoring 788
  - production (see Bremsstrahlung)
  - shielding 779-785

Gantry (see Medical accelerators)  
 Gas strippers (see Strippers)  
 Geiger-Müller counter 789  
 Genetic algorithm for optimization 83  
 Geophone 455  
 Giant dipole resonance 772  
 Gibbs phase rule 407  
 Glass 308, 464, 799  
 Glidcop® 442, 446, 673  
 Global accelerator network GAN 762  
 GMAW (see Welding)  
 Gradient undulator, size measurement 730  
 Gray 4, 767  
 Ground vibration (see Vibration)  
 GTAW (see Welding)  
 GTO 498, 536, 606  
 Gun (see ParticleSources, electron, Electron guns, Cathodes)  
 Gyromagnetic ratio 184  
 H<sup>-</sup> sources (see Particle Sources)  
 Hadronic cascade (see Showers)  
 Hadron-nucleus interaction (see Protons)  
 Haissinski equation 146  
 Hall probe (see Magnetic measurements)  
 Halo particles (see Beam halo)  
 Hamiltonian 89-93, 109 (see Maps)  
     canonical coordinates 101  
     crystalline beam 212  
     drift space 101  
     equation of motion, linear system 85  
     Lagrange invariant 85  
     hard edge of magnets 90  
     longitudinal 91, 117  
     near a single resonance 118  
     synchrobetatron 92, 93  
     transverse 91  
 Hardt condition (see Extraction)  
 Head-tail instability (see Collective instabilities)  
 Health physics (see Radiation protection)  
 Heat exchanger 417, 427  
 Heat shield (see Shielding, thermal)  
 Heat transfer 414-420  
     Fourier equations 415  
 Heavy Ion  
     colliders (see Colliders)  
     interactions 28, 282, 774-778  
     beam measurements 697-706  
 Helium 308, 407, 408  
 Health effects of ionizing radiation 768  
 Hereward feedback 707  
 Hexaborides 543, 544

High voltage  
     breakdown, dc (see Breakdown)  
     conditioning 513  
     cables 512  
     devices 51, 52, 513 (see Separators, electrostatic)  
     electrodes 512  
     electrodynamic accelerators 22-24  
     generators 22-24, 512  
     insulating core transformer (ICT) 23  
     insulator 511  
     technique 511-515  
 High voltage electrodynamic accelerator (see High voltage)  
 Higher order modes HOM 34, 55, 60, 127, 145, 431, 475-481, 494, 507, 618, 628-636, 664, 672, 674-681, 683-687, 750, 751 (see Couplers, Trapped modes)  
 coupler (see Couplers)  
 dampers 473-475, 672 (see Couplers)  
 Hilbert transformation 243  
 Higher order modes HOM  
 Hose instability 691  
 Hour glass effect 172, 176, 312 (see Beam-beam effects)  
 Hydrogen 308, 407, 408  
 Hysteresis (see Magnet)  
 ICT (see High voltage)  
 IGBT 498, 537, 539, 606  
 IGCT 501  
 Impedance, beam (see Collective instabilities, Wakefield)  
     approximate relation,  $Z_0^{\parallel}$  and  $Z_1^{\perp}$  243  
     beam position monitor 261, 710  
     Bethe small obstacles 245, 257  
     boundary conformal method 250  
     calculation, time domain 248-252  
     calculation, freq. domain 243-248  
     ceramic with metal coating 252, 517  
     closed pill-boxes 260  
     coherent synchrotron radiation (see Coherent synchrotron radiation)  
     collimator 259, 265, 340  
     corrugations 246  
     curvature effect 138, 168, 246, 260  
     definition and properties 242, 243  
     diffraction model 245, 256, 265  
     effective impedance 150-152, 262  
     electrostatic separators 614  
     explicit expressions 252-262  
     independence of beam direction 246

- kicker 260
- method of boundary perturbation 244
- method of effective boundary conditions 245
- method of integral equation 245
- moving window 250
- optical model 252
- $\omega^{-1/2}$  vs  $\omega^{-3/2}$  paradox 245
- periodic corrugations 258
- resistive pipe with dielectric coating 259
- resistive wall 148, 246, 253
- resistive wall, ac conductivity 253
- resistive wall, multilayer 246, 255
- resonator model 151, 243, 259
- shielding 246
- space charge 137-140, 253 (see Space charge)
- tapers 246
- trapped modes (see Trapped modes)
- wall roughness 246, 258
- Impedance, beam, determination
  - bench measurements 745-750
  - direct wakefield measurement 754-756
  - in situ beam measurements 750-752
- Impedance, cavity shunt 35, 478-481, 521, 549, 617-620, 647, 664-668, 670-690, 751
  - characteristic impedance 475, 593, 604
  - input 475, 593, 604
  - matching 475, 493
- Induced radioactivity 790
- Induction linac 24-27
  - core material 26
  - maximum current 26
- Inductively loaded accelerating cavities 667-670
- Inductive output tube IOT 659
- Industrial accelerator 27-31
  - material characterization 29, 30
  - material processing 28, 30
- Inflector (see Magnet, septum, pulsed)
- Injection 382-385
  - charge exchange 49, 383
  - charge exchange for FFAGs 383
  - continuous, or top up 273, 290
  - electron & positron injection 383
  - painting scheme 49, 383
  - pulsed sextupole 383
  - resonant injection 383
  - stacking with stochastic cooling 9, 200
- Inox 512
- Instabilities (see Collective instabilities)
- Instrumentation, beam (see individual measurements, Diagnostics, beam)
- Insulators
  - high voltage (see High voltage, insulator)
  - properties (see Dielectrics)
- Integrators
  - Euler method 109
  - high order symplectic 111
  - implicit mid-point 110
  - leap-frog 110, 249
  - Runge-Kutta 109
  - symplectic 109-111
  - symplectic tracking with differential algebra (see Differential algebra)
  - Yee algorithm 249
- Interaction mean free path 773
- Interferometer (see Laser interference methods)
- Intrabeam scattering 155-159
  - Coulomb log (see Coulomb log)
  - diffusion 158
  - emittance growth rates 156
  - in hadron colliders 158
  - stripping H<sup>-</sup> 159
  - temperature relaxation 156
  - Touschek effect (see Touschek effect)
- Invar 401
- Invariant
  - adiabatic 66, 67
  - Courant-Snyder 65, 72
  - kinematic invariant 104
  - nonlinear generalized Courant-Snyder 103
- Inverse Compton scattering 31
- Inverse free electron laser (IFEL) 31
- Ion beams 774
  - neutron production 774
- Ion implantation 28
- Ion induced desorption 282
- Ion pump (see Vacuum, pump)
- Ion sources (see Sources)
- Ion trapping 159, 160, 290
  - beam-ion instability 159-162
  - fast beam-ion instability 161, 162
  - trap potential, high voltage 513
- Ionization
  - avalanche 32
  - chamber 737, 790
  - collision 286
  - cooling (see Cooling, beam)
  - cross-sections 281, 286
  - damping (see Cooling)
  - losses 267, 293, 294
  - photo- 286
  - tunneling 286
- Ionization range of protons 773

- Ions
  - interactions 28, 282, 774, 776, 778
  - neutron generation 774
  - range-energy relations 774
- Iron
  - magnetic properties 471, 579, 596, 597
  - mechanical properties 310, 400-402
- Islands (see Phase space)
- Isochronous cyclotron (See Cyclotron)
- Isochronous rings 117, 375
- Isotope production (see Radioisotope)
- Jitter
  - beam 340, 341, 367, 370, 455, 459, 654, 763
  - compression 334
  - mechanical (see Vibration)
  - switching time 533-542, 606
- Johnsen effect (see Transition)
- Joule-Thompson valve 428
- KAM theorem, tori 103, 112
- Kapitza resistance 418, 678
- Kapton, polyimide 308, 795
- Keil-Schnell criterion (see Landau damping)
- Kerma 767
- Kick factor 264
- Kicker 602-607
  - stripline 603
  - traveling wave 602
- Kilpatrick criterion 53, 478, 663 (see Breakdown)
- Kinetic theory of gases 432 (see Conductance)
- Kinematic invariant (see Invariants)
- Kinematic relations 7
- Kink instability (see Beam-beam effects, linear colliders)
- Klein-Nishina cross-section 267, 275, 288
- Klystron 35, 651-653
  - multi-beam 652
  - amplifier systems 653-655
  - modulators 653-655
  - tubes 651-653
- Kohler's rule 466
- Kovar 447, 492, 519
- Kramers-Kronig relation 714
- KV distribution (see Envelope equation)
- Labyrinths (see Shielding)
- Lambertson septum (see Septum devices)
- Laminarity 133
  - emittance dominated, space charge dominated 133
- Landau damping 152-155
  - Boussard criterion 148, 154, 155
- dispersion relations 139, 153, 154
- Keil-Schnell criterion 148, 153, 154
- stability diagrams 153, 154, 744 (see Beam transfer functions)
- Laser accelerator
  - dielectric 31
  - photonic gap 31
  - plasma wakefield 31-33, 690-697
- Laser cooling (see Cooling, beam)
- Laser wire 726
- Laser interference methods
  - by gradient undulator 730
  - by double-slit interferometer 731
  - Michelson's stellar interferometer 731
  - transverse spot size 728
- Laslett tune shift 138-140, 253
- Lattice (see Optics, magnetic)
  - Chasman-Green, double bend achromat (DBA) 82, 117
  - collider storage rings 77-81
  - Combined function, FD cell 59, 79
  - double bend achromat (DBA) 82
  - design codes (See Computer codes, lattice)
  - dispersion suppressor 79
  - FODO cell 78, 79
  - global parameter grid scan 83
  - low-beta insertion 79
  - low emittance lattices 81-85
  - matching (see Optics, magnetic)
  - minimum emittance 82
  - multi-bend achromat (MBA) 82, 84
  - optimization techniques 83
  - separated function 59
  - simulation codes (See Computer codes, particle tracking simulations)
  - superbend 84
  - triple bend achromat (TBA) 82
- Lenses (see individual type)
- Lethal dose 769
- Lexan 308
- Lie algebra 99-105 (see Maps)
- Baker-Campbell-Hausdorff (BCH) theorem 100, 198
- Dragt-Finn factorization 99, 108
- Lie maps of individual elements 101
- Lie operator 99
  - spin motion 198
- Lienard-Wiechert potential 215
- Lifetime (see Beam lifetime)
- LIGA 689
- Light ion reactions 775
- Linac structures (see Accelerating structures)

- Linear accelerator
  - electrons 34-36
  - protons 36-40
  - equations of motion 133-137, 141, 321, 557
  - parameters scaling with frequency 665
  - beam-loss detection 328, 737
  - SLAC parameters 664
  - superconducting linac 326-329
- Linear accelerator operation 321-329, 667
- Linear collider (see Collider)
- Linear energy transfer 767, 790
- Linear transformer driver (LTD) (see Pulsed high voltage devices)
- Liouville theorem 144, 387
- Liquefier (He) 425-429
- Liquid critical, triple point, values 406
- LIS 570
- Lithium lens 9, 567, 621-624
- Litz wire 11, 582
- Livingston chart 12, 40
- Log ratio detector 711
- Longitudinal beam distribution measurement 713-719 (see Bunch distribution and Streak camera)
- Longitudinal emittance measurement (see Energy spread, Longitudinal beam distribution and Bunch distribution)
- Longitudinal mode coupling instability (see Collective instabilities)
- Longitudinal motion (oscillations) 66, 67
  - linac 141, 321, 557
  - rf bucket (see Rf bucket)
  - synchrotron 58, 67
    - strong rf focusing 390
- Lorentz detuning 48, 328, 649, 654, 677
- Lorentz force 6
- Lorentz stripping (see Stripping)
- Lorentz transformation 8, 216
- Loss factors 263-265, 676, 746, 753
  - collimator 265
  - diffraction model 265
  - resistive wall 265
  - resonator impedance 264
- Lossy materials (rf) 473-475
- Low- $\alpha$  lattice 117
- Low beta insertion (see Lattice)
- Low emittance light sources 81-85 (see Emittance)
- Low level rf (LLRF) 649-651
- Lucite 308, 464, 795
- Luminosity 80, 169, 311-313, 338
  - average luminosity 313-318
- crossing angle 312, 339
- dynamic focusing (see Beam-beam effect, dynamic beta)
- enhancement factor 176, 312, 338
- factories 311, 313
- hour glass effect (see Hour glass effect)
- integrated, optimization 314
- linear collider 178
- monochromatization (see Monochromatization)
- monitoring 273
- optimal coupling 311
- reductions 312
- transparency 311
- Luminosity lifetime (see Beam lifetime)
- Lyapunov exponent (see Dynamic aperture)
- Macor 464
- Magnet 577-614
  - ac 581, 582 (see Betatron)
  - cam mover (See Cam mover)
  - coil construction 581
  - core construction 579
  - current density 581
  - eddy currents (see Eddy currents)
  - ends 578-581
  - field properties, analytic 585
  - hysteresis 596
  - keystoning 581, 599
  - kicker (see Kicker)
  - laminations (see Magnet, core construction)
  - measurement 523-529
  - multipole 577, 582, 607-614 (see Multipole expansion)
  - overhang formulae 585
  - permanent 471-473, 607-614
  - permanent quadrupole 607-614
  - persistent currents 591, 592
  - pole shaping (shims) 578, 579, 585
  - power dissipation 588, 596
  - pre-stress 591, 600
  - pulsed 602-607 (see Kicker)
  - quality control 582
  - resistive 577-584
  - saturation effects 578, 584, 585, 590, 595
  - septum (see Septum devices)
  - sextupole 578, 613
  - solenoid (see Solenoid)
  - superconducting- $\cos \theta$  588-595
  - superconducting-high field 599-602
  - superconducting-superferric 595-599
  - support (See Supports, equipment)
- Magnetic materials 471-475, 579, 597, 607, 667

- Magnetic rigidity (see Rigidity)
- Magnetic shielding 503, 678, 707
- Magnetic measurements 523-533
  - accelerator magnets 523-529
  - insertion devices 529-533
  - Hall probe 528, 532, 533
  - NMR probe 382, 528
- Magnetically insulated transmission line, MITL (see Transmission lines)
- Magnetoresistivity (see Kohler's rule)
- Magnetron discharge 518
- Magnetron source 564
- Maps 97-105 (see Hamiltonian, Optics, Lattice)
  - Cremona map 98, 104
  - from field data 75-77, 102
  - hard edge fringe field 74, 76
  - Hénon map 98
  - Lie map 99-105 (see Lie algebra)
  - nonlinear transport maps 97, 101
  - normal form 97, 102, 107
  - reversed 107
  - standard map 98
  - Taylor map 97-99 (see Differential algebra)
- Marx generator (see Pulsed high-voltage devices)
- Matching, impedance (see Impedance, transmission line)
- Matching, optics (see Optics, magnetic)
- Material properties (see also individual materials and properties)
  - atomic and nuclear (values) 308, 310
  - electrical (magnetic)(values) 308, 310 463, 475, 795
  - mechanical-thermal (values) 399-408, 463, 464, 795
  - properties of some thermoset and thermoplastic materials 795 (see Thermal conductivity, Thermal expansion, Elastic modulus, Specific heat)
- Maxwell equations 6
  - boundary conditions 6
  - grid equations 248-251
- Medical accelerators 27, 40-44, 385
  - brachytherapy (see Brachytherapy)
  - gantry 21, 41
  - intensity-modulated radiation therapy (IMRT) 41
  - neutron therapy 43 (see Boron neutron capture therapy)
- Metglas 26
- Metalization 447
- MEVVA 568
- Micro-dosimetric principles 789
- Microtron 329
- Microwave instability (see Collective instabilities)
- Mid-plane symmetry, magnets 72
- MIG (GMAW) (see Welding)
- Millimeter wave accelerators 687-690
- Minimum propagation zone (see Quench)
- Mismatch (see Emittance dilution)
- Möbius accelerator 88
- Model independent analysis MIA 367-373
- Modes
  - bunch oscillation 150-152
  - cavity 478-481, 670
  - fast wave, slow wave 153, 354
  - HOM (see Higher order modes)
  - spectral density 262
  - transmission lines, waveguides 477, 478
- Modeling, operations
  - linac 321, 367-370
  - storage ring 357-360, 370-373
- Modulator (rf) 35, 651-655
- Møller scattering 273, 756
- Moment matrix 72, 86
  - equation of motion 70, 72
  - nonlinear 104
- Momentum compaction factor 66 (see Slip factor)
  - $\alpha_P < 0$  lattice (see Slip factor)
- FODO cell 78, 117
- nonlinear 116, 149
- Momentum dispersion (see Dispersion function)
- Momentum measurement, ions 699
- Monochromatization, storage ring collider 313
- Monte carlo codes 776
  - FLUKA 776
  - GEANT4 776
  - MARS15 777
  - MCNPX 778
  - PHITS 778
  - MOSFET 498, 537
  - Moyer model 780
  - Multi-bend achromat (see Lattice)
  - Multi-bunch crossing instability (see Beam-beam effects, linear colliders)
  - Multicusp source 564
  - Multi-moded rf components 486-490
  - Multipacting 35, 163, 494-496 (see Electron-cloud effect)
  - Multiple scattering 267, 271, 361
  - Multipole expansion 93, 523-526, 589
    - cylinder analysis 75
  - Muon 785
    - production 293, 777

- range-energy relation 268, 293, 294, 785
- shielding (see Shielding)
- Muon collider 44-47, 88, 297-300
  - ionization cooling (see Cooling, ionization)
  - muon storage ring neutrino factory 46
  - backgrounds (see Backgrounds)
- Mylar 308, 795
- Natural background radiation 769
- Nb (see Niobium)
- NbTi 404, 465-471
- NEG pump (see Vacuum pump)
- Negative mass instability (see Collective instabilities, negative mass)
- Neodymium iron boron NdFeB 290, 472
- Neutrino factory, muon based 46
- Neutron
  - activation reactions 790-793
  - compact source 48
  - fluence to dose conversion coefficients 768
  - monitoring 786
  - shielding 773, 783 (see Shielding)
  - spallation source (see Spallation neutron source)
  - target 50
  - yield (see Radioactivity)
- Neutron angular distribution 775
- Neutron dosimetry and Spectrometry 786
- Neutron production 36, 293, 769, 772, 773
  - compact sources 48
  - spallation source 47-51
  - target 50
- Nichrome 465
- Niobium 401-425, 465-471, 678
  - electropolishing 422
  - fabrication 420-425
- Nitrogen 308, 310, 407, 408, 425
- NMR probe (see Magnetic measurements)
- Noise
  - BPM 369
  - dipole 361
  - rf 361, 646, 649
- Nonlinear dynamics 89 (see Maps, Dynamic aperture)
  - experiments 118-121
- Nonlinear resonance (see Resonance)
- Normal form (see Maps)
- Nuclear cascade 293, 776
- Nuclear collision length 308, 310
- Nuclear interaction length 308, 310
  - values
- Numerical integration of motion 109-111
- Nusselt number 416
- Nutriino factory (See Muon collider)
- Nylon 308, 795
- Octupoles (see Magnet, Optics, Operational considerations)
- Oide limit 277, 278
- Oil, transformer 465
- Operational considerations
  - circular machines 346-387
  - linacs 124-127, 321-346
  - rf setpoint for proton linacs 38
- Optics, magnetic 71-89
  - matching 77
- Optical model (see Impedance, beam)
- Optical stochastic cooling (see Stochastic cooling)
- Orbit distortion 346
  - $7 \times 7$  method 70
  - dependence on beam current 753
  - recirculative in ERL 129
- Orbit correction 350-352 (see Feedback)
  - global 350, 351
  - local bump 351, 352
  - response matrix 357, 624
  - singular value decomposition, SVD 350, 625
- Orbital eigenanalysis 68-71
- Organ absorbed dose 767
- Outgassing
  - ion-induced 278-283
  - photon-induced 278-283
  - pressure bump instability 159
  - thermal 278 (see Thermal desorption)
- Ovvovoltage factor 222
- Painting (see Injection)
- Pair production
  - cross channel process 277
  - electron 179, 293, 555, 558
  - linear collider 179
- Panofsky quadrupole 596
- Panofsky-Wenzel theorem 201, 242
- Parametric radiation 240
- Parasitic beam-beam effects 180-183
  - long range effects 181-183
  - pretzel scheme 181
  - separation criterion 183
  - separation scheme 80
- Parasitic loss 263-265 (see Loss factors)
- Paraxial ray approximation 134
- Parasitic modes (see HOM)
- Particle induced x-ray emission (PIXE) 29
- partition numbers, synchrotron radiation (see Radiation damping)

- Paschen curve 518, 534
- Pauli matrices 184
- Pelletron 17
- Penning discharge 435
- Penning gauge 440
- Penning source 564, 568
- Peralumin 512
- Periodic table 469
- Permanent magnet 471-473, 607-614 (see Magnetic materials, Magnet)
- Permeability, magnetic 394, 411, 471, 578, 579, 603, 667-670
- Permeability effects (see Magnet, saturation effects)
- Permittivity (see Dielectric constant)
- Persistent currents (see Magnet, Eddy currents)
- Personal dose equivalent 768
- Perveance (see Electron guns)
- PFN, PFL 602-607 (See also Modulator)
- Phase displacement acceleration (see Rf gymnastics)
- Phase mismatch 361
- Phase oscillation (see Longitudinal motion)
- Phase slippage factor (see Slip factor)
- Phase space 65-71
  - chaos (see Dynamic aperture)
  - ellipse 72
  - islands, island tune 114, 118
  - separatrix 94, 384, 385
  - surface of section 118
- Phase stability
  - principle 58, 67
  - boundary 67, 141
- Phase velocity  $v_p$  34, 477
- Phasor diagram 129
- Photodesorption 278-283
- Photoelectron instability (see Electron-cloud effect)
- photoelectron yield 283, 284 (also see Quantum efficiency)
- Photon (see Gamma radiation, Bremsstrahlung)
- Photon dosimetry and spectrometry 788
- Photon flux 81, 225
- Photon Transmission 781
- Photonic gap structure 31
- Photoneutrons 772
- Pickup electrodes, PUE (see Current, Position)
- Pierce gun 550
- PIG 568
- Pinch enhancement, luminosity (see Luminosity, Beam-beam, disruption)
- Piwinski angle (see Beam-beam effects, storage ring colliders)
- Plasma accelerator 32, 690-697
  - beat-wave 692
  - blowout regime 33, 692
  - channel guiding 693
  - laser wakefield 33, 691
  - optical guiding 693
  - plasma wakefield 32, 691
  - wave-breaking limit 32, 690
- Plasma frequency 32, 266, 690
- Plasma lens 32, 694,
- Plasma wavelength 691
- Plexiglas 308
- Pockels cell 547
- Poisson's ratio 414, 463, 464, 556
- Poisson bracket 99, 110, 199
- Polarimeters 756-760
  - Compton, electrons 757
  - Møller, electrons 756, 757
  - protons 758
- Polarization
  - beam-beam depolarization, linear collider 178
  - depolarization resonances (see Resonances, depolarization)
  - electron energy calibration 193, 382
  - measurement (see Polarimeters)
  - of synchrotron radiation 225, 226
- Polarization, radiative 191-195
  - Baier-Katkov spin flip 191
  - Baier-Katkov-Strakovenko equation 192
  - depolarization time 192, 195, 197
  - Derbenev-Kondratenko-Mane formula 192, 194
  - equilibrium polarization 191, 192, 195
  - polarization time 191, 192
  - radiative depolarization 192
  - SLIM formalism 194, 195
  - spin orbit  $\hat{n}_0, \partial\hat{n}/\partial\gamma$  (see Spin closed orbit)
  - Sokolov-Ternov effect 191
- Polarized particle sources (see Sources)
- Polarized protons, acceleration of 187-191
- Pole shape (see Magnet)
- Polyethylene 308, 464, 795, 796
- Polyphase circuits 496
- Polystyrene 308, 464
- Ponderomotive force 691, 693
- Population exposure (see Radiation dose)
- Position measurement BPM 347, 709-713
- Positron emission tomography PET 43
- Positron sources (see Sources)

- Potential well distortion 145-147
- CSR 228
- Power converter 498-501
- Power supplies (see Power converter)
- Poynting vector 6, 215
- Prandtl number 416
- Preinjector (see Electron preinjector, Electron guns)
- Proton
  - elastic p-N scattering (see Scattering cross-section)
  - interactions 269-271, 295, 773
  - monitoring 788
  - muon production by 293, 777, 785
  - neutron production by 773
  - range-energy relation 773
  - removal mean free path 293, 773
- Proton beams 773
  - neutron yield 773
- PTFE (see Teflon)
- Pulse compression (rf) 481-486
  - active switching 485, 486
  - binary pulse 484
  - SLED 482, 483
  - SLED-II 483
- Pulsed high-voltage devices 26, 51, 52 (see High voltage)
  - linear transformer device (LTD) 52
  - Marx generator 51
  - pulse transformer 52
  - voltage and current fluctuations 52
- Pulsed magnet (see Magnet)
- Pumps and pumping (see Vacuum)
- Q*, cavity 479, 482, 493, 522, 523
- Quadrupoles (see Magnet, Optics, Operational considerations)
- Quantum efficiency QE 164, 544, 546,
- Quantum excitation 70, 221
- Quantum lifetime 222
- Quarter wave transformer
  - rf 475
  - optics, positron collection 557
- Quartz 463
- Quasi-isochronous rings (see Isochronous rings)
- Quench
  - detection 594
  - minimum propagation zone 594
  - propagation velocity 594
  - protection 593,
- Rad 767
- R matrices (see Transport matrices)
- Radiation (see Heat transfer)
- Radiation attenuation (see Shielding)
- Radiation damage 50, 793-802
  - at cryogenic temperatures 796
  - ceramics 799
  - crystal 290, 638
  - drift chamber 290
  - electronics 290, 797-799
  - glass 799
  - metals 801
  - optical fibers 800
  - organic materials 793-796
  - permanent magnetic materials 290
  - scintillator 290, 799
  - tests 794
  - vitreous materials 799-801
- Radiation damage thresholds 793
  - ceramics 799
  - cryogenic temperatures 796
  - metals 801
  - optical fibers 800
  - radiation damage tests 794
  - scintillators 799
  - vitreous materials 799
- Radiation damping 70, 220
  - partition numbers 221
- Radiation dose
  - calculations (see Shielding)
  - lethal 769
  - measurement (see Radiation measurement)
  - natural 769
  - rate 769, 770, 779, 791, 794, 800
  - standards 769
- Radiation index 794
- Radiation integrals 82, 220
  - generalized in coupled system 71
  - FODO cell 78
- Radiation length  $X_0$  238, 308, 310
- Radiation measurements 786-790
  - gamma (see Gamma, monitoring)
  - muon (see Muon, monitoring)
  - neutron (see Neutron, monitoring)
  - proton (see Proton, monitoring)
- Radiation protection 767-793
  - quantities and terms 767
  - standards (see Radiation dose)
- Radiation protection principles
  - Physical quantities 767
- Radiation shielding (see Shielding)

- Radiation sources 769-776
  - gamma 770
  - muon 785
  - neutron 771,
- Radiation terms (see Radiation protection)
- Radiation therapy (see Medical accelerators)
- Radiation transport (see Shielding)
- Radiation transport and Shielding 779
  - Monte carlo “hybrid” method 780
  - Semi-empirical Models 779
- Radiative energy loss for heavy particles 293-296
- Radiative polarization (see Polarization, radiative)
- Radioactivity 790-793
  - calculation of 776, 790
  - exposures from 769
  - half life, values 791
  - in the environment 769
  - radionuclides 791
  - neutron yield 790-793
  - units 4, 767
- Radio frequency quadrupole, RFQ 52-55
  - beam dynamics 53
  - Kapchinskiy-Teplyakov (KT) procedure 53
  - rf structure 54
  - space charge (see Space charge)
  - superconducting 54
  - tuning 39, 54
- Radiography 25, 44, 52
- Radioisotope 43 (see Radioactivity)
- Radionuclides (see Radioactivity)
- Range (see  $dE/dx$ , Shielding, radiation attenuation)
- Rayleigh length (range) 319
- Recirculating linac (See Energy recovery linac)
- Recombination chamber 790
- Recommended dose limits 769
- Reflection coefficient, transmission line 475
- Refractive index 308, 310, 463, 464
  - values
- Refrigerators, LHe (see Cryogen, refrigerators)
- Relative Biological Effectiveness RBE 769
- Relativistic klystron (see Two-beam accelerator)
- Relativity 8
- Rem 767
- Rem counters 786
- Residual gas analyzer (see Vacuum, RGA)
- Resistive wall instability (see Collective instabilities)
- Resistivity (see Conductivity)
- Resonances
  - coupling 68, 354
  - cyclotron 16
  - difference and sum resonances 68, 95, 354
  - driving term 103, 119
  - experiments 118-121
  - extraction (see Extraction, slow)
  - in cyclotron 16
  - in FFAG 19, 21
  - nonlinear 94
  - in proton linac 143
  - synchrobetatron coupling (see Synchrobetatron effects)
  - synchrobetatron (satellite) resonances 94-97
  - synchrobetatron resonance above and below transition 95
  - nonlinear resonance width, strength 94, 107
- Resonances, depolarizing 186-188, 193, 195, 196
- equation of motion 186
- Froissart-Stora equation (see Froissart-Stora equation)
- intrinsic and imperfection 186, 187
- snake resonances 187
- Rexolite (see Polystyrene)
- Reynolds number 416
- Rf breakdown (see Breakdown)
- Rf bucket 67, 141 (see Phase space)
  - area 67
  - barrier bucket (see Rf gymnastics)
  - boundaries (see Phase stability)
  - height 141, 222
  - mismatch 361
  - nonlinear momentum compaction 117
  - separatrix 141 (see Phase stability)
- Rf cavity (See Accelerating structures, Cavity)
- Rf distribution
  - DLLS
- Rf frequency bands 481
- Rf guns, electron 134, 327, 548, 550
  - charge and wavelength scaling 137
- Rf gymnastics 376-381
  - adiabaticity 376
  - barrier bucket 9, 380, 381
  - batch compression 380
  - using broadband RF systems 380
  - bunch compression 334-337, 377, 685
  - bunch rotation 8, 377
  - coalescing 379
  - controlled blow-up 376
  - debunching 378
  - merging and splitting 378
  - phase displacement acceleration 380
  - slip stacking 379

- stacking (see Injection)
- strong rf focusing 390
- Rf low level (LLRF) 649-651
- Rf noise (see Noise)
- Rf separators (see Separators)
- Rf structure (see Accelerating structures)
- Rf system design 645, 649
- Rhodotron 55, 56
- Richardson's law (see Electron guns)
- Rigidity  $B\beta$  1, 73, 217
  - cyclotron 14
  - electrostatic 704
- Ripple effects (see Dynamic aperture, Extraction, Eddy currents)
- Robinson instability (see Collective instabilities)
- Robinson theorem, sum rule 70, 87, 221
- Rose shims (see Magnet, pole shaping)
- RRR (residual resistivity ratio) 405, 420, 466, 678
  - (see individual materials)
- Rutherford backscattering, RBS 29
- $S$ -matrix, rf 745 (see Cavity, measurements)
- Sacherer formulae 150-152
- Samarium cobalt SmCo 290, 472
- Sapphire 400, 463, 465, 522
- SASE (see FEL)
- Saturation, magnet (see Magnet)
- Scattering cross-section
  - antiproton production 8
  - Bhabha (see Bhabha scattering)
  - Bremsstrahlung (see Bremsstrahlung)
  - Compton (see Compton scattering)
  - Coulomb (see Coulomb scattering)
  - Klein-Nishina (see Klein-Nishina cross-section)
  - Møller (see Møller scattering)
  - $pN$  269, 295, 296
  - $pp, p\bar{p}$  270, 295, 296
  - Thomson 267
- Schottky effect 543
- Schottky spectra 739-741, 742-745 (see Stochastic cooling)
  - longitudinal unbunched 742
  - longitudinal bunched 742
  - momentum measurement 742
  - transverse unbunched 740
  - transverse bunched 740
- Schwinger critical field 177
- Scintillators 308, 799
- Secondary emission (see Electron-cloud effect, Multipacting)
  - yield 164, 284, 285, 463, 495, 515
  - effective SEY  $\delta_{\text{eff}}$  164
- Seismometer 455
- Self amplified spontaneous emission SASE (see Free electron laser)
- Separators
  - electrostatic 614-616 (see Breakdown, High voltage)
  - rf 617
- Separatrix (see Phase space)
- Septum devices 639-641
  - Lambertson 641
  - electrostatic 513, 639
- Sextupoles (see Magnet, Optics, Operational considerations)
- Shear strength 403
- Shell effects, stripping (see Stripping)
- Shielding, magnetic (see Magnetic shielding)
- Shielding materials 784
- Shielding, radiation 779-785
  - attenuation 771, 774, 779-784, 786
  - calculations 776, 779
  - coherent synchrotron radiation (see Coherent synchrotron radiation)
  - computer codes 776
  - electron accelerators 781
  - labyrinths, ducts and doors 782, 783
  - materials 784
  - muons 785
  - proton accelerators 779 (see Moyer model)
  - radiation transmission 780
  - removal mean free path 773
  - tunnels 771, 780
- Shielding, thermal 419, 430
- Showers 293, 555, 771, 772
  - hadronic 293
- Shunt impedance (see Impedance, cavity shunt)
- Siberian snake 185, 187-191, 641
  - partial snake 185, 188, 642
  - types 1 & 2 185
- Sievert 767
- Silver 465
- Single bunch instabilities (see Collective instabilities)
  - linacs 123-125
  - circular accelerators 147-150
- Singular value decomposition, SVD 367, 350, 625 (see Orbit correction, Model independent analysis)
- Size measurement, beam 713-732
  - emittance, transverse and longitudinal 699, 702, 713-719, 721, 723-726, 728, 730
  - laser wire (see Laser wire)

- laser interference (see Laser interference methods)
- longitudinal distribution, coherent radiation 713
- longitudinal distribution, electro-optical 714
- streak camera (see Streak camera)
- Skin depth 148, 253, 475, 587, 671
  - anomalous 475
- Skyshine 785
- SLED (see Pulse compression)
- Slip factor 91, 116-118 (see Momentum compaction factor)
  - $\eta < 0$  lattice 148
  - nonlinear 116
- Sloan-Lawrence accelerator 53
- Slow extraction (see Extraction)
- Smith-Purcell radiation 241
- Sokolov-Ternov effect (see Polarization, radiative)
  - spectrum 177
- Solenoid (see Transport, Coupling, particle motion)
  - compensation 355-357
  - in ebis ion source 568
  - high field 601
  - in linear-collider final focus 340
  - permanent 612
  - positron collection 557
- Sources, particle
  - antiproton 8-10, 565-567
  - deuterons 562
  - electrons 543-555
  - electrons, polarized 546
  - H<sup>-</sup> 563-565
  - ion (see Strippers)
  - ion, polarized 561, 562
  - neutron spallation (see Spallation source)
  - optically pumped 561
  - positrons 555-558
  - positrons, polarized 558-561
- Sources of radiation 769-776
- Space charge
  - beam potential 91, 159 (see Beam-beam, storage ring colliders)
  - centrifugal 138, 168
  - circular accelerators 137, 138
  - cyclotron 16
  - compensation, guns+transport 133-137, 136
  - compensation, storage rings 394
  - curvature effect (see Impedance, curvature effect)
  - guns 133-137
- impedance (see Impedance, beam)
- induction linacs 26
- inductive inserts 394
- limiting current, DC gun 134
- proton linac 49, 142
- RFQ 53
- transport line 133-137
- transverse s.c. compensation 395
- Spallation neutron source 47-51 (see Neutron production)
- Spark gap 26, 51, 533, 539, 606
- Specific heat 310, 401, 408, 463, 464, 598, 616, 623
- Specific energy 775
- Spectral density, beam (see Modes, bunch oscillation, Beam transfer functions)
- Spin closed-orbit  $\hat{n}_0$  184, 185, 191
  - Derbenev-Kondratenko vector  $\hat{n}$  108, 192, 194
  - spin invariant field 192
  - spin-orbit coupling vector  $\partial\hat{n}/\partial\delta$  192, 195
- Spin echo (see Echo)
- Spin flip & spin flippers 186-188
- Spin matching 193, 195-198
  - beam-beam 198
  - computer codes 194, 195
  - transparency 196
- Spin motion (see Thomas-BMT equation, Stern-Gerlach force, Polarization, Siberian snake)
  - figure-8 ring 185
  - helical dipole 185
  - Lie algebra 198, 199
  - Siberian snake (see Siberian snake)
  - SLIM formalism (see Polarization, radiative)
  - Thomas-BMT equation (see Thomas-BMT equation)
- Spin precession 183, 638 (see Thomas-BMT equation)
- Spin rotator 185, 196, 643
- Spin transparency (see Spin matching)
- Spin tune (see Tune)
- Spinor algebra 184
- Spoke cavity (see Accelerating structures)
- Sputter ion pump, SIP (see Vacuum, pump)
- Stability diagrams (see Landau damping)
- Stacking (see Injection)
- Stainless steel 400-403 (see Vacuum chamber)
- Stanton number 416
- Steel (see Iron)
- Stern-Gerlach force 198

- Stochastic cooling 8, 9, 199-202, 397
  - bandwidth 199
  - bunched beam 200
  - coherent electron cooling 202
  - cooling rates 199
  - filter 201
  - kicker 201
  - longitudinal 199
  - longitudinal stacking 200
  - optical stochastic cooling 201
  - operational considerations 397
  - pickup 201
  - potential theorem 201
  - transverse 199
  - unbunched beam 199
- Stopping power (see  $dE/dx$ )
- Storage ring 56-58
  - collider (see Collider)
  - ultimate (see Ultimate storage ring)
- Strange attractor (see Dynamic aperture)
- Streak camera 723-726
- Stress-strain relations 404 (see individual materials and composites)
- Stress relaxation (see Creep)
- Strippers (see Stripping)
- Stripping 571-577
  - blackbody stripping 49
  - charge exchange cross section 571
  - charge exchange injection 36, 383, 576
  - charge strippers 571-574
  - cyclotron extraction 16
  - deuteron (d, n) 775
  - efficiency 574
  - foils 571, 574
  - intrabeam scattering of  $H^-$  (see Intrabeam scattering)
  - laser 576
  - Lorentz stripping 15, 44, 49, 576
  - most probable charge state 704
  - shell effects 572
- Strong focusing (see Focusing)
- Superbend lattice (see Lattice)
- Super insulation (see Shielding, thermal)
- Superconductors
  - BCS surface resistance 678
  - electrical properties 465-471, 678
  - fabrication of rf structures 420-425
  - high temperature 467
  - materials 468, 469
  - mechanical-thermal properties 404-406
  - wire/cable 466, 467
- Superfluid heat transfer 418
- Supports, equipment 451-453
- Surface resistance 475 (see Skin depth)
- Survey (see Alignment)
- Switches
  - high power (see individual devices) 533-542
  - DC 533-540
  - rf 540-542
  - optical 542
  - pseudospark 535, 539
- Symbols, definitions 1-3
- Symplectic conjugate 353
- Symplectic integrators (see Integrators)
- Symplecticity 69, 85, 99, 109-111
- Synchrobetatron effects
  - beam-beam with crossing angle (see Beam-beam effects, storage ring colliders)
  - coupling 68, 92, 93
  - resonance (see Resonance)
- Synchrocyclotron (see Cyclotron)
- Synchronous phase 58, 140
- Synchrotron 58, 59
- Synchrotron oscillations (see Longitudinal motion)
- Synchrotron radiation 215-242
  - bending magnet 217-220
  - coherent (see Coherent synchrotron radiation)
  - critical photon energy 218, 225
  - final focus 340 (see Oide limit)
  - low- $\alpha$  lattice (see Low- $\alpha$  lattice)
  - magnet edge 237
  - point charge 215
  - polarization 225, 226
  - power 215, 218
  - radiation damping (see Radiation damping)
  - radiation integrals (see Radiation integrals)
  - shielding 219
  - shielding for backgrounds (see Backgrounds)
  - storage rings 220-223
  - undulator and wiggler (see Undulator, Wiggle)
- Synchrotron tune (see Tune)
  - with nonlinear momentum compaction 117
- Tandem accelerator 17, 23
- Taylor maps (see Maps)
- TBA two-beam accelerator (see Two-beam accelerator)
- TBA triple bend achromat (see Lattice)
- Teflon (PTFE) 308, 401, 402, 464, 795
- TEM cavities 681
- TEM transmission lines 477, 478, 602-607
  - power ratings 475

- Temperature coefficient of resistance 465, 466
- Tensile strength 403, 463, 464, 795
- Tetrode amplifiers 655-659
  - beam loading correction 133
- Therapy (see Radiation therapy)
- Thermal conductivity 415
  - values 310, 400, 405, 408, 463, 464, 795
- Thermal diffusion (see Heat transfer)
  - diffusivity 453
  - in earth 453
- Thermal desorption 278-283
- Thermal distortion 452
- Thermal expansion, values 401, 406, 463, 464, 795
- Thermal photons 277, 290, 341
- Thermal radiation 418
- Thermoacoustic oscillations 429
- Thermomagnetic properties 466
- Thick targets 770, 771
- Thin film 516-521
  - beam impedance (see Impedance, beam)
  - field penetration 517
  - power loss 517
- Thin-target bremsstrahlung 771
- Thomas-BMT equation 178, 183, 186, 191, 192, 198
- Thomson cross-section 267
- Thyatron 26, 52, 534, 535, 606
- Thyristor 498, 536, 537, 539, 606
- Ti (see Titanium and alloys)
- Ti sublimation pump (see Vacuum, pump)
- TIG (GTAW) (see Welding)
- Timing systems 763 (see Control system)
- TiSP (see Vacuum, pump)
- Titanium and alloys 308, 400-406, 512, 515-521, 560 (see Vacuum chamber)
- Titanium nitride 518, 519
- Touschek effect 157, 272, 289 (see intrabeam scattering)
  - lifetime 158
- Track etched detectors 788
- Transfer lines (cryogenics) 429-431
- Transfer map (see Maps)
- Transformer ratio (see Wakefield accelerator)
- Transit time factor 140, 521
- Transition
  - crossing 373-376
  - emittance growth, impedance 374
  - gamma 67, 116
  - $\gamma_t$  jump 375
  - energy 58, 67
- Johnsen effect 374
- Umst tter effect 374
- Transition radiation 226, 719
- Transmission lines 475-478, 593, 602-607
  - pulsed high voltage 51
  - TEM (see TEM transmission lines)
- Transport matrices 71-75 (see Maps)
  - beam rotation 75
  - bending magnet 73
  - drift 73
  - edge focusing, bend (see Maps)
  - nonlinear (see Maps)
  - pole face rotation 74
  - quadrupole, thick and thin 73
  - principle planes, bend 74
  - R* matrix 72, 717
  - skew quadrupole 75
  - solenoid 74, 352, 353
  - solenoid and quad, superimposed 353
- Trapping, particles 208, 214
- Trapped modes 265
- Traveling wave (see Accelerating structures)
- Transverse emittance measurement (see Size measurement)
- Transverse mode coupling instability (see Collective instabilities)
- Triple bend achromat TBA (see Lattice)
- Triple point 406
  - values 407
- Triplet, quadrupoles 80
- Truncated power series algebra, TPSA (see Differential algebra)
- Tune
  - amplitude dependence, synchrotron tune 67
  - anharmonicity 98, 103
  - betatron 66, 103
  - coupled system 87
  - depression 143
  - diffusion (see Diffusion)
  - eigen-tunes 69
  - island tune (see Phase space)
  - measurement 349 (see Schottky spectra)
  - momentum and amplitude dependence, betatron tunes 93, 103, 108, 119
  - modulation (see Dynamic aperture)
  - real time measurement & control 362-365
  - shift, octupole 93
  - shift, potential well (see Potential well distortion)
  - shift, quadrupole error 93, 346
  - shift, space charge (see Space charge)
  - sidebands (see Schottky spectra)

- spin 184, 185, 192-194, 199  
 synchrotron 67

Tungsten and alloys 308, 512, 556, 561, 566

Tuner (see Tuning)

Tuning  
 drift tube linacs 663  
 FFAG 19  
 multicell cavities 678  
 single cell cavities 673, 686

Two-beam accelerator TBA 60-62  
 compact linear accelerator scheme 60  
 relativistic klystron 60

Two-particle models (see Collective instabilities)

Umst tter effect (see Transition)

Ultimate storage ring 84

Undulator (see Wiggler, Free electron laser)  
 field 223, 609  
 gradient undulator 730  
 polarized e<sup>+</sup> production (see Sources)  
 radiation 81, 223-226, 229, 320, 558  
 undulator parameter 223, 229

Units 4, 5  
 Gaussian-MKSA conversions 432

Vacuum 431-448  
 chamber 440-445  
 ceramic chamber (see Ceramic vacuum chamber)  
 conductance (see Conductance)  
 crotches 446  
 gauges 439, 440  
 desorption by residual gas ions 281  
 heavy-ion induced desorption 282  
 performance 436  
 photon-induced gas desorption 279  
 pumps, pumping 435-439  
 pumping speed 433-435  
 RGA 440  
 stability limit 281  
 transition materials 444  
 UHV materials 442

Vacuum pumps  
 distributed 438  
 NEG 437  
 sputter ion pump, SIP 435  
 Ti sublimation pump TiSP 437

Van Citter-Zernike's theorem 732

Van de Graaff 17

Vector analysis 8

Velocity bunching 136

Velocity measurement 699-702

Vibration

control 452, 459-461  
 emittance growth, linacs 324-329  
 ground 453-459  
 ground wave 457-461  
 instruments 454, 455  
 luminosity reduction 456

Viscosity, values 408

Vitreous materials 799-801 (see Glass)

Vlasov equation 144, 145  
 linearization 145

Voltage standing wave ratio, VSWR 475, 522 (see S-matrix, rf)

Wakefield (see Impedance, beam, Collective instabilities)  
 causality 243  
 for a Gaussian beam 261  
 in linacs 34, 264, 340  
 indirect integration 251  
 limiting values 243  
 properties 242, 243  
 wake-induced beam tail 716-719

Wakefield accelerator 32  
 dielectric 32  
 plasma (see Plasma accelerator)  
 laser 33, 687, 691 (see Laser accelerator)  
 transformer ratio 32, 691, 693

Wake free steering 325

Wake function 123, 242 (see Impedance)  
 disk loaded waveguide, empirical fit 257, 322

Wake potential 128, 242

Wall current monitor 707

Water 308, 407, 408, 416-420

Wave-breaking limit (see Plasma accelerators)

Waveguide 475-478  
 circular 478  
 coaxial 478  
 rectangular 477  
 US standards 477

Weak focusing (see Focusing)

Wehnelt cylinder 547

Welding  
 DEPI 444  
 dissimilar metals 444  
 electron beam 421, 444  
 friction 444  
 GMAW (MIG) 443  
 GTAW (TIG) 444

Wien filter 185

Wiggler (see Undulator)  
 critical photon energy 225  
 damping wiggler 84, 222  
 radiation 223-225, 226

Wigner distribution (see Brightness)  
Windows, rf  
    coating 515, 516  
    design 491-493  
Work function 134  
    Schottky 134  
    values 513  
  
X-ray fluorescence 288  
  
Yee algorithm (see Integrators)  
Yield strength 402  
Yoke (see Magnet, core construction)  
Young's modulus (see Elastic modulus)  
  
Z-pinch 52

## Errata and Additions to Handbook, 2nd Edition

### January 18, 2016

p.170, left column, Dynamic emittance

Original text

New text A more complete expression for the dynamic emittance was computed in:

- A.V.Otboyev, E.A.Perevedentsev, PRST-AB 2, 104401 (1999)

p.261, second row from the top, longitudinal impedance and wake of a BPM consisting of two strip lines

Original text  $Z_0^{\parallel} = 2Z_c \left[ \frac{\phi_0}{2\pi} \right]^2 [2 \sin^2 kL - i \sin 2kL]$ ,  $W'_0 = 2Z_c c \left[ \frac{\phi_0}{2\pi} \right]^2 [\delta(z) - \delta(z+2L)]$

New text  $Z_0^{\parallel} = Z_c \left[ \frac{\phi_0}{2\pi} \right]^2 [2 \sin^2 kL - i \sin 2kL]$ ,  $W'_0 = Z_c c \left[ \frac{\phi_0}{2\pi} \right]^2 [\delta(z) - \delta(z+2L)]$

p.572, right column, Lifetimes of carbon foils

Original text  $k_{\text{foil}} = 0.018$  for foils produced by vapor deposition

New text  $k_{\text{foil}} = 0.0018$  for foils produced by vapor deposition

p.576, right column, References

Original text

New text Remove Ref.[1]. Add two additional references:

- L.R. Scherk, Can. J. Phys. 57, 558 (1979)
- P.B. Keating et al., Phys. Rev. A 52, 4547 (1995)