

CERN 94-01  
26 January 1994  
Vol. I

ORGANISATION EUROPÉENNE POUR LA RECHERCHE NUCLÉAIRE

**CERN** EUROPEAN ORGANIZATION FOR NUCLEAR RESEARCH

**CAS** CERN ACCELERATOR SCHOOL

FIFTH GENERAL ACCELERATOR PHYSICS COURSE

University of Jyväskylä, Finland

7–18 September 1992

PROCEEDINGS

Editor: S. Turner  
Vol. I

GENEVA  
1994

© Copyright CERN, Genève, 1994

Propriété littéraire et scientifique réservée pour tous les pays du monde. Ce document ne peut être reproduit ou traduit en tout ou en partie sans l'autorisation écrite du Directeur général du CERN, titulaire du droit d'auteur. Dans les cas appropriés, et s'il s'agit d'utiliser le document à des fins non commerciales, cette autorisation sera volontiers accordée.

Le CERN ne revendique pas la propriété des inventions brevetables et dessins ou modèles susceptibles de dépôt qui pourraient être décrits dans le présent document; ceux-ci peuvent être librement utilisés par les instituts de recherche, les industriels et autres intéressés. Cependant, le CERN se réserve le droit de s'opposer à toute revendication qu'un usager pourrait faire de la propriété scientifique ou industrielle de toute invention et tout dessin ou modèle décrits dans le présent document.

Literary and scientific copyrights reserved in all countries of the world. This report, or any part of it, may not be reprinted or translated without written permission of the copyright holder, the Director-General of CERN. However, permission will be freely granted for appropriate non-commercial use.

If any patentable invention or registerable design is described in the report, CERN makes no claim to property rights in it but offers it for the free use of research institutions, manufacturers and others. CERN, however, may oppose any attempt by a user to claim any proprietary or patent rights in such inventions or designs as may be described in the present document.

ISSN 0007-8328  
ISBN 92-9083-057-3

CERN 94-01  
26 January 1994  
Vol. I

ORGANISATION EUROPÉENNE POUR LA RECHERCHE NUCLÉAIRE  
**CERN** EUROPEAN ORGANIZATION FOR NUCLEAR RESEARCH

 CERN ACCELERATOR SCHOOL

FIFTH GENERAL ACCELERATOR PHYSICS COURSE

University of Jyväskylä, Finland  
7–18 September 1992

PROCEEDINGS  
Editor: S. Turner  
Vol. I

GENEVA  
1994



## **ABSTRACT**

The fifth CERN Accelerator School (CAS) basic course on General Accelerator Physics was given at the University of Jyväskylä, Finland, from 7 to 18 September 1992. Its syllabus was based on the previous similar courses held at Gif-sur-Yvette in 1984, Aarhus 1986, Salamanca 1988 and Jülich 1990, and whose proceedings were published as CERN Reports 85-19, 87-10, 89-05 and 91-04, respectively. However, certain topics were treated in a different way, improved or extended, while new subjects were introduced. As far as the proceedings of this school are concerned the opportunity was taken not only to include the lectures presented but also to select and revise the most appropriate chapters from the previous similar schools. In this way the present volumes constitute a rather complete introduction to all aspects of the design and construction of particle accelerators, including optics, emittance, luminosity, longitudinal and transverse beam dynamics, insertions, chromaticity, transfer lines, resonances, accelerating structures, tune shifts, coasting beams, lifetime, synchrotron radiation, radiation damping, beam-beam effects, diagnostics, cooling, ion and positron sources, RF and vacuum systems, injection and extraction, conventional, permanent and superconducting magnets, cyclotrons, RF linear accelerators, microtrons, as well as applications of particle accelerators (including therapy) and the history of accelerators.

**SEFT**



# CERN ACCELERATOR SCHOOL

RESEARCH INSTITUTE FOR HIGH ENERGY  
PHYSICS, FINLAND

UNIVERSITY OF JYVÄSKYLÄ

The Institute will organize a course on

## GENERAL ACCELERATOR PHYSICS

at Jyväskylä, Finland, 13-18 September 1992

This course will mainly be suitable for researchers in accelerator laboratories, university departments and companies specialising in equipment for particle accelerators. It requires knowledge up to first-degree level of physics, mathematics and engineering.

For further information and application forms, please contact:

Mrs Sven Warrberg  
CERN Accelerator School  
CH-1211 Geneva 23  
Switzerland

TELEPHONE

+41 22 767 80 00

FAX NUMBER

+41 22 767 81 86

Administrative  
Programme Committee

Local Organizing Committee

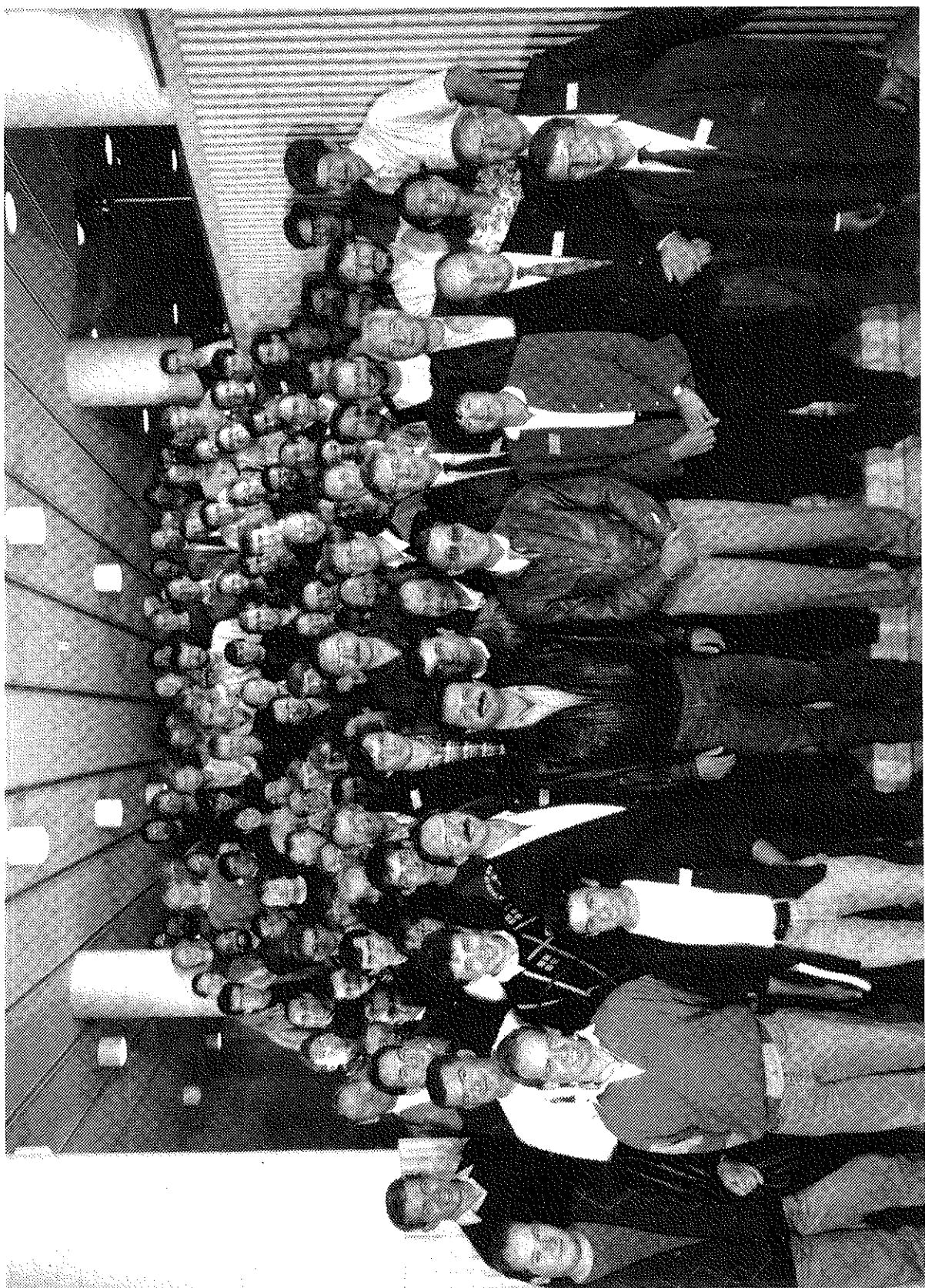
Chairman  
Professor P. Hervet  
University of Jyväskylä

Secretary  
Mrs S. Warrberg  
University of Jyväskylä

**CLOSING DATE FOR APPLICATIONS: 15 JUNE 1992**

Sponsored by: The Finnish Ministry of Education

PROGRAMME FOR BASIC COURSE ON GENERAL ACCELERATOR PHYSICS, Jyväskylä, Finland, 7-18 September, 1992



## FOREWORD

CERN's Accelerator School was established in 1983 with the mission to preserve and disseminate the knowledge accumulated at CERN and elsewhere on particle accelerators and storage rings of all kinds. This is being carried out principally by means of a biennial programme of basic and advanced two-week courses on general accelerator physics aiming to bridge the gap between the level of knowledge attained with a science or engineering degree and that required for starting accelerator research work. In addition, topical courses are organised jointly with the US Particle Accelerator School, while specialised courses are staged as the need or opportunity arises and in conjunction with appropriate bodies or organizations. This policy has led to the vigorous programme of courses and their proceedings shown in the table overleaf.

The present proceedings are for the fifth basic course on general accelerator physics held at the University of Jyväskylä, Finland in September 1992. The course followed closely the earlier ones at Gif-sur-Yvette, Aarhus, Salamanca and Jülich but with more emphasis on cyclotrons on this occasion. However, to avoid readers having to refer to all the previous proceedings in order to have a complete coverage of the basic course topics, the opportunity was taken to select and revise the most appropriate chapters from previous proceedings and to bring them together in the present volumes.

Without the very serious help and encouragement of the organisations and individuals involved it would not have been possible to hold the Jyväskylä school, nor to publish its proceedings. The particularly generous financial and organizational support of the Finnish Ministry of Education, SEFT, JYFL, the University and Municipality of Jyväskylä and MAP Medical Technologies Oy are most gratefully acknowledged while the time and effort devoted to the school by many individuals, but especially Risto Orava, Markus Nordberg and Pauli Heikkinen, deserve all our thanks. As ever, the continuing support of the CERN Management, the School's Advisory, Programme and Local Organization Committees were most important. Very special thanks must go to all the lecturers at the school who gave so much of their time to prepare, present and write-up their topics. The major efforts made by the many authors to bring together into one set of proceedings the most appropriate topics of all the previous basic courses are much appreciated. Many CERN services made invaluable contributions to the success of this school while the efficiency and friendliness of the management and staff at the hotel Alba made our stay in Jyväskylä most enjoyable. Finally we thank the students who came from all over the world to attend the course. That they were so numerous, enthusiastic and friendly made the efforts made on their behalf so worthwhile.

S. Turner, Editor

## LIST OF CAS COURSES AND THEIR PROCEEDINGS

Year	General Accelerator Physics Courses	Topical Courses	Specialised Courses and Workshops
1983	-	-	Antiprotons for colliding beam facilities CERN 84-15 (1984)
1984	Basic CERN 85-19 (1985)	-	Generation of high fields (ECFA and INFN Workshop) ECFA 85/91, CERN 85-07 (1985)
1985	Advanced CERN 87-03 (1987)	Nonlinear dynamics aspects of particle accelerators Lecture Notes in Physics 247 (Springer-Verlag, Berlin, 1986)	-
1986	Basic CERN 87-10 (1987)	Frontiers of particle beams Lecture Notes in Physics 296 (Springer-Verlag, Berlin, Heidelberg, New York, 1988)	Applied Geodesy for particle accelerators CERN 87-01 (1987) Applied Geodesy Lecture Notes in Earth Sciences 12 (Springer-Verlag, Berlin, Heidelberg, New York, 1987)
1987	Advanced CERN 89-01 (1989)	-	New developments in particle acceleration techniques (ECFA et al. Workshop) CERN 87-11, ECFA 87/110
1988	Basic CERN 89-05 (1989)	Frontiers of particle beams: Observation, diagnosis and correction Lecture Notes in Physics 343 (Springer-Verlag, Berlin, Heidelberg, New York, 1989)	Superconductivity in particle accelerators CERN 89-04 (1989)
1989	Advanced CERN 90-04 (1990)	-	Synchrotron radiation and free-electron lasers CERN 90-03 (1990)
1990	Basic CERN 91-04 (1991)	Frontiers of particle beams: Intensity limitations Lecture Notes in Physics 400 (Springer-Verlag, Berlin, Heidelberg, New York, 1992)	Power converters for particle accelerators CERN 90-07 (1990)
1991	Advanced CERN 92-01 (1992)	-	RF engineering for particle accelerators CERN 92-03 (1992), 2 Vols.
1992	Basic Present volumes	Frontiers of particle beams: Factories with $e^+e^-$ rings Lecture Notes in Physics 425 (Springer-Verlag, Berlin, Heidelberg, New York, 1994)	Magnetic measurements and alignment CERN 92-05 (1992)
1993	Advanced (to be published)	-	RF engineering for particle accelerators Repeat of 1991 course
1994	Basic (to be published)	Frontiers of accelerator technology (to be published)	Cyclotrons, linacs and their applications (to be published)

## CONTENTS

	Page no.
<b><u>Volume I:</u></b>	
Foreword	vii
<b><i>P.J. Bryant</i></b>	
<b>A brief history and review of accelerators</b>	<b>1</b>
Introduction	1
Historical roots	1
The main development	9
The current situation in high-energy particle physics accelerators	11
Conclusion	14
<b><i>J. Rossbach, P. Schmüser</i></b>	
<b>Basic course on accelerator optics</b>	<b>17</b>
Introduction	17
Accelerator magnets	21
Particle motion in a circular accelerator	34
Betatron oscillations	46
Motion of particles with momentum deviation	69
Comparison of strong and weak focusing	79
<b><i>J. Buon</i></b>	
<b>Beam phase space and emittance</b>	<b>89</b>
Introduction	89
Summary of classical and elementary results	90
Phase portraits	95
Beam emittance: A statistical point of view	103
<b><i>K. Potter</i></b>	
<b>Luminosity measurements and calculations</b>	<b>117</b>
Introduction	117
Definition of luminosity	117
Luminosity of a single-ring collider	118
The RMS as a measure of beam height	119
Luminosity with coasting beams and finite crossing angle	121
Measurement of the luminosity	124
The Van der Meer method of luminosity measurement	125
Vertical beam displacements at the ISR	127
Concluding remarks	127

<b>E. Wilson</b>	
<b>Transverse beam dynamics</b>	<b>131</b>
Introduction	131
Liouville's theorem	132
A simplified treatment of betatron motion	139
The $Q$ values	141
Closed-orbit distortion	143
Gradient errors	151
The working diagram	153
Chromaticity	154
Conclusions	157
<b>P.J. Bryant</b>	
<b>Insertions</b>	<b>159</b>
Introduction	159
Matching	160
Dispersion suppressors embedded in a FODO lattice	167
Low- $\beta$ insertions	171
Extraction and injection	173
Collimation insertions	174
More exotic insertions	187
<b>S. Guiducci</b>	
<b>Chromaticity</b>	<b>191</b>
Introduction	191
Quadrupole	192
Sextupole	194
Chromaticity correction	195
General bending magnet	197
End-field effects	201
<b>P.J. Bryant</b>	
<b>A simple theory for weak betatron coupling</b>	<b>207</b>
Introduction	207
Coupling in uniform skew quadrupole and axial fields	208
Observations	214
Results of the exact analysis	216
<b>P.J. Bryant</b>	
<b>Beam transfer lines</b>	<b>219</b>
Distinctions between transfer lines and periodic circular machines	219
Orbit correction in transfer lines	221
Matching transfer lines	225
Emittance and mismatch measurement in a dispersion-free region	225
Small misalignments and field ripple errors in dipoles and quadrupoles	227
Emittance blow-up due to thin windows in transfer lines	231
Emittance dilution from betatron mismatch	232
Emittance exchange insertion	235
<b>E. Wilson</b>	
<b>Non-linearities and resonances</b>	<b>239</b>
Introduction	239
Multipole fields	241
Second-order resonance	242
The third-integer resonance	246
General numerology of resonances	249
Slow extraction using the third-order resonance	249
Landau damping with octupoles	250

<b>J. Le Duff</b>	
<b>Dynamics and acceleration in linear structures</b>	<b>253</b>
Basic methods of linear acceleration	253
Fundamental parameters of accelerating structures	259
Energy gain in linear accelerating structures	273
Particle dynamics in linear accelerators	277
<b>J. Le Duff</b>	
<b>Longitudinal beam dynamics in circular accelerators</b>	<b>289</b>
Acceleration by time varying fields	289
Dispersion effects due to the guide field in a synchrotron	296
Synchrotron oscillation in adiabatic limit	299
Adiabatic damping of synchrotron oscillations	306
Trapping, matching, accumulating and accelerating processes	309
<b>T. Risselada</b>	
<b>Gamma transition jump schemes</b>	<b>313</b>
Introduction	313
Effect of quadrupoles on $\gamma_t$	316
Zero tune shift and non-zero $\Delta(\gamma)$	318
Cells, superperiods and families	321
Conclusion	325
<b>A. Hofmann</b>	
<b>Tune shifts from self-fields and images</b>	<b>329</b>
Introduction	329
Direct incoherent tune shifts	330
Effects of the walls on the incoherent tune shift	336
Coherent tune shifts	340
Comments	344
Practical examples	346
<b>J.-L. Laclare</b>	
<b>Coasting beam longitudinal coherent instabilities</b>	<b>349</b>
Introduction	349
Single-particle motion	351
Longitudinal signal of a single particle	353
Distribution function	356
Electromagnetic field induced by the beam	360
Negative mass instability	363
Introduction of the longitudinal coupling impedance $Z_{\parallel}(\omega)$	364
Longitudinal coupling impedance $Z_{\parallel}(\omega)$ of an accelerator ring	366
Vlasov's equation and dispersion relation	371
Monochromatic beam	375
Coasting beam with momentum spread	378
Landau damping by momentum spread	380
Limits of the theory	382
<b>J.-L. Laclare</b>	
<b>Coasting beam transverse coherent instabilities</b>	<b>385</b>
Introduction	385
Single-particle transverse motion	386
Transverse signal of a single particle	387
Distribution function	390
Total beam signal	390
Definition of transverse coupling impedance	391
Transverse coupling impedance $Z_{\parallel}(\omega)$ of an accelerator ring	393
Dispersion relation for coherent motion	395
Beam without tune spread	397
Landau damping by momentum spread	401
Landau damping by amplitude dependent tune	405

<b>A. Wrulich</b>	
<b>Single-beam lifetime</b>	<b>409</b>
Introduction	409
Lifetime due to quantum fluctuation and radiation damping	412
Lifetime for statistical fluctuations without damping	414
Lifetime due to beam-gas scattering	416
Lifetime due to Touschek scattering	424
Intrabeam scattering	428
Lifetime due to resonance crossing	432
Lifetime measurement	433
<b>R.P. Walker</b>	
<b>Synchrotron radiation</b>	<b>437</b>
Introduction	437
Basic properties	438
Spectral and angular properties	442
Photon distribution	449
Synchrotron radiation aspects in electron accelerator design	451
Synchrotron radiation sources	453
Synchrotron radiation from protons	454
<b>R.P. Walker</b>	
<b>Radiation damping</b>	<b>461</b>
Introduction	461
Energy oscillations	462
Betatron oscillations	466
Damping partition and the Robinson theorem	470
Radiation damping aspects in various lattice designs	471
Modification of damping rates	475
Measurement of damping rates	477
<b>R.P. Walker</b>	
<b>Quantum excitation and equilibrium beam properties</b>	<b>481</b>
Introduction	481
Energy oscillations	481
Betatron oscillations	486
Synchrotron radiation integrals	489
Quantum lifetime	490
Low emittance lattices	493
Changes in beam properties due to insertion devices	494
<b>Volume II:</b>	
<b>H. Mais, C. Mari</b>	
<b>Introduction to beam-beam effects</b>	<b>499</b>
Introduction	499
Basic facts	500
Linear beam-beam models	504
Experimental facts and results for lepton and hadron colliders	510
Theoretical tools and methods	515
Summary and conclusions	520
<b>Y. Baconnier, A. Poncet, P.F. Tavares</b>	
<b>Neutralisation of accelerator beams by ionisation of the residual gas</b>	<b>525</b>
Neutralisation of a beam: a simple description	525
The ionisation process	527
The ion or electron motion	530
A few examples of ion or electron motion	537

Bunched beams	539
Clearing electrodes	541
The limit of accumulation	541
The effects of neutralisation	541
Diagnostics and phenomenology	549
<b>H. Koziol</b>	
<b>Beam diagnostics for accelerators</b>	<b>565</b>
Introduction	565
Description of diagnostic devices	567
Concluding remarks	590
<b>S.P. Møller</b>	
<b>Cooling techniques</b>	<b>601</b>
Introduction	601
Stochastic cooling	602
Electron cooling	607
Laser cooling	613
Other cooling methods	615
Conclusion and comparisons	616
<b>N. Angert</b>	
<b>Ion sources</b>	<b>619</b>
Introduction	619
Principles	619
Ion sources for positive ions	624
Ion sources for negative ions	633
Beam formation	637
<b>R. Chehab</b>	
<b>Positron sources</b>	<b>643</b>
Physical processes associated with positron production	643
Presentation of some positron sources	655
Positron collection: The matching system	661
Emittance transformation and preservation	674
Comparison of positron sources	674
Summary and conclusions	675
<b>M. Puglisi</b>	
<b>Conventional RF system design</b>	<b>679</b>
Introduction	679
The accelerating gap	679
The drift tube	680
Cavity resonators	682
The cylindrical cavity	684
TEM cavities	686
$\lambda/2$ cavity	688
Re-entrant cavity	689
Common RF accelerating structures	691
Coupling to the cavities	693
Shunt impedance	695
RF power amplifiers	696
RF generators	696
<b>A.G. Mathewson</b>	
<b>Vacuum system design</b>	<b>717</b>
Introduction	717
Basic formulae	717
Conductance	718
Monolayer	719

Pumping systems	720
Cleaning methods	720
Thermal outgassing	722
Bakeout	723
Synchrotron radiation induced gas desorption	724
Simple machine	726
Proton storage rings	727
Cold proton storage rings	729
<b>G.H. Rees</b>	
<b>Injection</b>	<b>731</b>
Introduction	731
Single-turn injection	732
Multi-turn injection	734
H <sup>-</sup> charge-exchange injection	737
Injection from a cyclotron into a synchrotron	742
Novel injection schemes	742
<b>G.H. Rees</b>	
<b>Extraction</b>	<b>745</b>
Introduction	745
Fast extraction	746
Slow extraction	747
Septum units	751
<b>S. Wolff</b>	
<b>Superconducting accelerator magnet design</b>	<b>755</b>
Introduction	755
Superconducting magnet configurations	755
Field analysis	757
Examples of existing magnet designs	765
Design details	767
Test results at HERA mass production magnets	783
Conclusions	788
<b>H. Lengeler</b>	
<b>Modern technologies in RF superconductivity</b>	<b>791</b>
Limitations of SC cavity performances	792
Nb material	794
Surface diagnostics, inspection and repair methods	794
Shaping and welding of SC cavities	797
Surface treatments	799
Cavities coated with thin superconducting layers	801
<b>P. Heikkinen</b>	
<b>Cyclotrons</b>	<b>805</b>
Introduction	805
Operation principle	805
Isochronous cyclotron	807
Cyclotron hardware	815
<b>P. Heikkinen</b>	
<b>Injection and extraction for cyclotrons</b>	<b>819</b>
Introduction to injection systems	819
Neutral beam injection	820
Axial injection	820
Radial injection	826
Longitudinal matching	830
Introduction to extraction schemes	830
Extraction systems for cyclotrons	831

Stability of vertical motion in extraction	836
Extraction elements	836
Conclusions	838
<b>O. Barbalat</b>	
<b>Applications of particle accelerators</b>	<b>841</b>
Introduction and overview	841
Research applications	842
Element analysis	846
Medicine	846
Industrial processing	848
Power engineering	851
Conclusion	852
<b>W. Pohlit</b>	
<b>Accelerators for therapy</b>	<b>855</b>
Introduction	855
Development of present tumor therapy	855
Developments for future radiation therapy of tumors	860
<b>N. Marks</b>	
<b>Conventional magnets - I</b>	<b>867</b>
Introduction	867
Magneto-static theory	867
Practical aspects of magnet design	875
<b>N. Marks</b>	
<b>Conventional Magnets - II</b>	<b>891</b>
Introduction to AC effects	891
Low frequency systems	892
High frequency magnets	901
Conclusion	910
<b>M. Weiss</b>	
<b>Introduction to RF linear accelerators</b>	<b>913</b>
Introduction	913
Electromagnetic waves and cavities	914
Wave equation and slowing down of waves	917
Travelling and standing wave linear accelerators	921
Electromagnetic forces and their action	928
Phase stability and acceptance of an accelerator	932
Handling of intense beams	939
Various structures of linear accelerators	941
Conclusion	951
<b>P. Suortti</b>	
<b>Photon beamlines and monochromators</b>	<b>955</b>
Introduction	955
X-ray optics	956
Constructions	964
Experimental stations	968
<b>P. Lidbjörk</b>	
<b>Microtrons</b>	<b>971</b>
Introduction	971
The circular or classical microtron	972
The racetrack microtron	977

<b>T. Meinander</b>	
<b>Generation of magnetic fields for accelerators         with permanent magnets</b>	<b>983</b>
Introduction	983
Permanent magnet materials	983
Calculation of permanent magnet fields	985
Electromagnets versus permanent magnets	986
Multipole magnets	990
Insertion devices	994
<b>H. Lengeler</b>	
<b>Nuclear waste transmutation using high-intensity         proton linear accelerators</b>	<b>999</b>
Introduction	999
The nuclear waste problem	999
Earlier transmutation studies	1001
A new Los Alamos proposal	1001
State of the art of high-power proton and deuteron linear accelerators	1004
Two design aspects of a high-intensity proton-linac	1006
A possible layout	1009
Superconducting (S.C.) cavities	1012
Conclusion	1016
List of participants	1021

# A BRIEF HISTORY AND REVIEW OF ACCELERATORS

*P.J. Bryant*  
CERN, Geneva, Switzerland

## ABSTRACT

The history of accelerators is traced from three separate roots, through a rapid development to the present day. The well-known Livingston chart is used to illustrate how spectacular this development has been with, on average, an increase of one and a half orders of magnitude in energy per decade, since the early thirties. Several present-day accelerators are reviewed along with plans and hopes for the future.

## 1 . INTRODUCTION

High-energy physics research has always been the driving force behind the development of particle accelerators. They started life in physics research laboratories in glass envelopes sealed with varnish and putty with shining electrodes and frequent discharges, but they have long since outgrown this environment to become large-scale facilities offering services to large communities. Although the particle physics community is still the main group, they have been joined by others of whom the synchrotron light users are the largest and fastest growing. There is also an increasing interest in radiation therapy in the medical world and industry has been a long-time user of ion implantation and many other applications. Consequently accelerators now constitute a field of activity in their own right with professional physicists and engineers dedicated to their study, construction and operation.

This paper will describe the early history of accelerators, review the important milestones in their development up to the present day and take a preview of future plans and hopes.

## 2 . HISTORICAL ROOTS

The early history of accelerators can be traced from three separate roots. Each root is based on an idea for a different acceleration mechanism and all three originated in the twenties.

### 2.1 The main "History Line"

The first root to be described is generally taken as the principal "history line", since it was the logical consequence of the vigorous physics research programme in progress at the turn of the century. Indeed, particle physics research has always been the driving force behind accelerator development and it is therefore very natural to also consider high-energy physics as the birth place.

The main events along this "history line" are listed in Table 1. The line is started at the end of the last century to show the natural progression through atomic physics to nuclear physics and the inevitable need for higher energy and higher intensity "atomic projectiles" than those provided by natural radioactive sources. In this context, the particle accelerator was a planned development and it fulfilled its goal of performing the first man-controlled splitting of the atom. It was Ernest Rutherford, in the early twenties, who realised this need, but the electrostatic machines, then available, were far from reaching the necessary voltage and for a few years there was no advance. Suddenly, the situation changed in 1928, when Gurney and Gamov independently predicted tunnelling [1] and it appeared that an energy of 500 keV might just suffice to split the atom. This seemed technologically feasible to Rutherford and he immediately encouraged Cockcroft and Walton to start designing a 500 kV particle accelerator.

Four years later in 1932, they split the lithium atom with 400 keV protons. This was the first fully man-controlled splitting of the atom [2] which earned them the Nobel prize in 1951.

**Table 1**  
Main "History Line"

1895	Lenard. Electron scattering on gases (Nobel Prize).	< 100 keV electrons. Wimshurst-type machines.
1913	Franck and Hertz excited electron shells by electron bombardment.	
1906	Rutherford bombards mica sheet with natural alphas and develops the theory of atomic scattering.	Natural alpha particles of several MeV
1911	Rutherford publishes theory of atomic structure.	
1919	Rutherford induces a nuclear reaction with natural alphas.	
	... Rutherford believes he needs a source of many MeV to continue research on the nucleus. This is far beyond the electrostatic machines then existing, but ...	
1928	<b>Gamov predicts tunnelling and</b> perhaps 500 keV would suffice ...	
1928	Cockcroft & Walton start designing an 800 kV generator encouraged by Rutherford.	
1932	Generator reaches 700 kV and Cockcroft & Walton split lithium atom with only 400 keV protons. They received the Nobel Prize in 1951.	

Figure 1(a) shows the original apparatus, which is now kept in the Science Museum, London. The top electrode contains the proton source and was held at 400 kV, the intermediate drift tube at 200 kV and final drift tube and target at earth potential. This structure can be seen inside the evacuated glass tube in Fig. 1 above the curtained booth in which the experimenter sat while watching the evidence of nuclear disintegrations on a scintillation screen. The voltage generator, Fig. 1(b), was at the limit of the in-house technology available to Cockcroft and Walton and the design voltage of 800 kV was never reached due to a persistent spark discharge which occurred at just over 700 kV. However, the famous atom-splitting experiment was carried out at 400 kV, well within the capabilities of the apparatus. The **Cockcroft Walton generator**, as it became known, was widely used for many years after as the input stage (up to 800 kV) for larger accelerators, since it could deliver a high current.

At about the same time Van de Graaff, an American who was in Oxford as a Rhodes scholar, invented an electrostatic generator for nuclear physics research and later in Princeton, he built his first machine, which reached a potential of 1.5 MV [3]. It took some time to develop the acceleration tube and this type of machine was not used for physics research until well after the atom had been split in 1932. The principle of this type of generator is shown in Fig. 2.

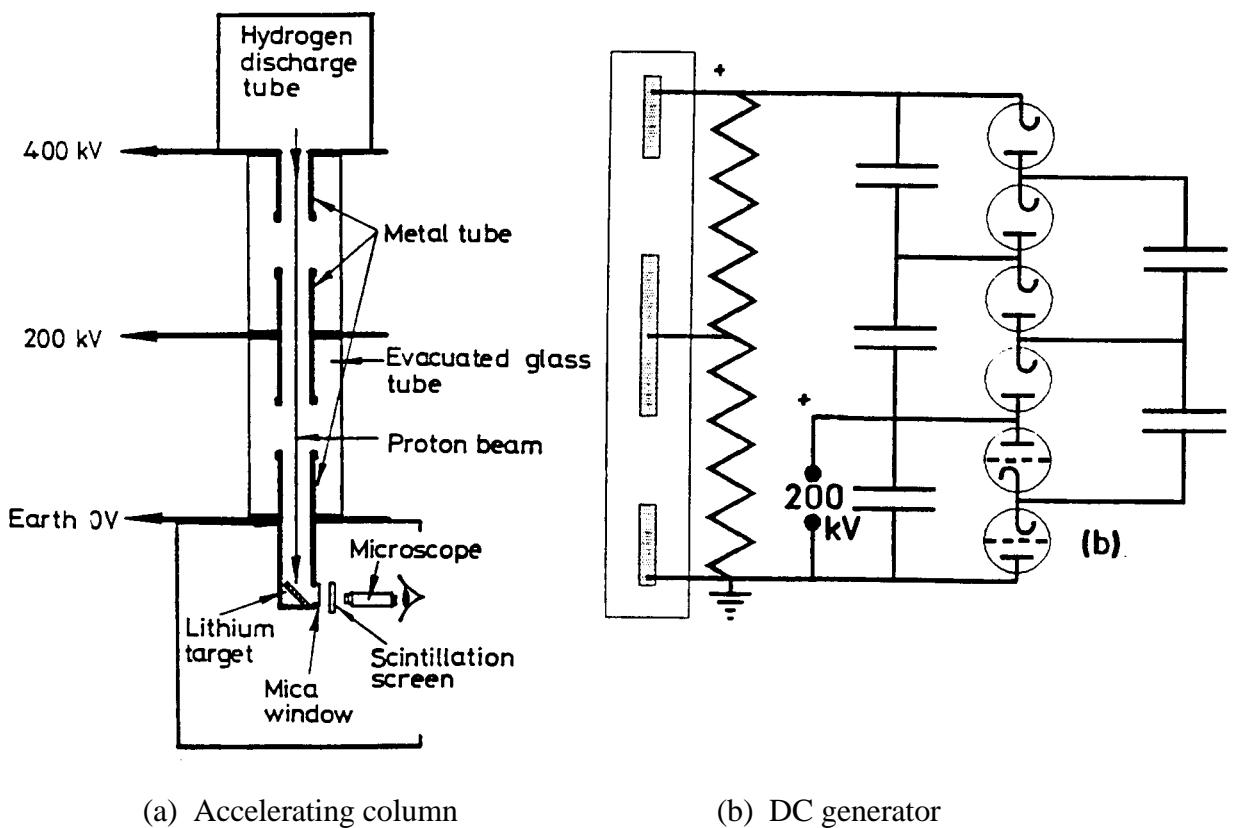
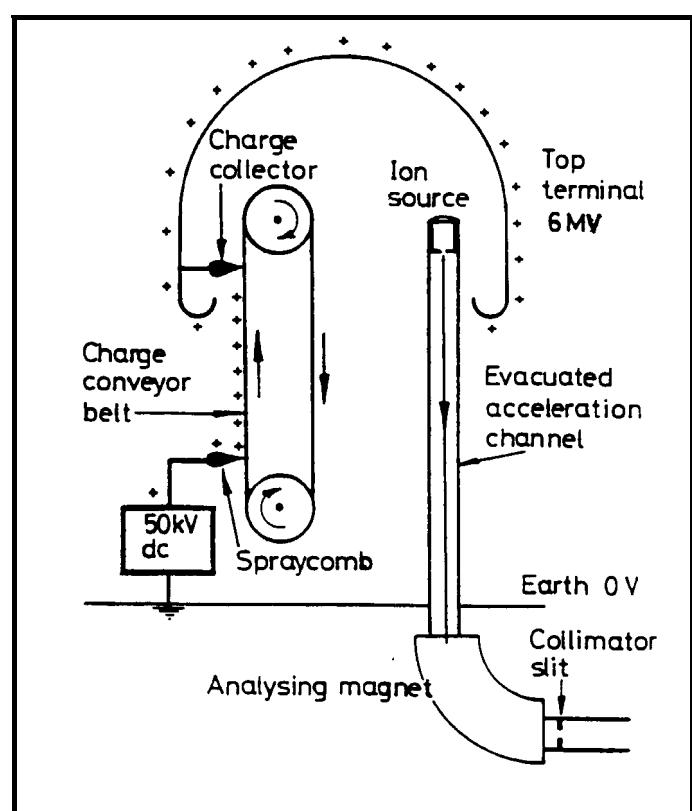


Fig. 1 Cockcroft and Walton's apparatus for splitting the lithium nucleus

Fig. 2 Van de Graaff electrostatic generator



Two new features appeared in later versions of the **Van de Graaff generator**. Firstly, the sparking threshold was raised by putting the electrode system and accelerating tube in a high-pressure tank containing dry nitrogen, or Freon, at 9-10 atmospheres, which enables operation typically up to 10 MV. The second was a later development, which has the special name of the **Tandem** accelerator (see Fig. 3).

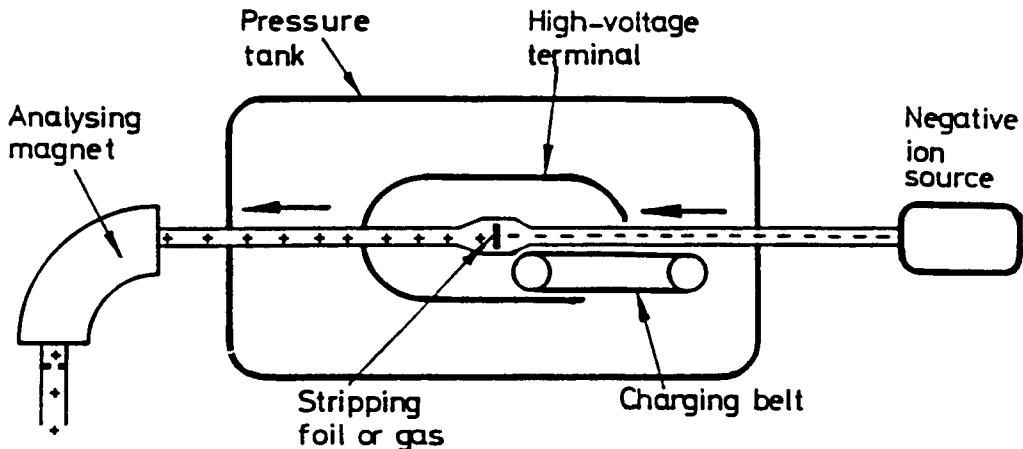


Fig. 3 Two-stage Tandem accelerator

The new feature in the Tandem accelerator was to use the accelerating voltage twice over. First an extra electron is attached to the neutral atoms to create negative ions. In recent years, a great deal of development has been done and it is now possible to obtain negative ion sources for almost all elements. The negative ion beam is injected at ground potential into the Tandem and accelerated up to the high-voltage terminal where it passes through a thin foil which strips at least two electrons from each negative ion converting them to positive ions. They are then accelerated a second time back to earth potential. The Van de Graaff generator and the Tandem provide beams of stable energy and small energy spread, but they are unable to provide as high currents as the Cockcroft-Walton generator

The highest energy Tandem is at Oak Ridge National Laboratory and routinely operates with 24.5 MV on the central terminal. However, development is not at a standstill and there is a project (the Vivitron) underway at Strasbourg to build a Tandem operating at 35 MV.

## 2.2 The second "History Line"

The direct-voltage accelerators were the first to be exploited for nuclear physics research, but they were limited to the maximum voltage that could be generated in the system (except for the astute double use of the applied voltage in the Tandem). This limitation was too restrictive for the requirements of high-energy physics and an alternative was needed.

In fact, an alternative had already been proposed in 1924 in Sweden by Ising [4]. He planned to repeatedly apply the same voltage to the particle using alternating fields and his invention was to become the underlying principle of all of today's ultra-high-energy accelerators. This is known as **resonant acceleration**. The main events along this "history line", starting with Ising, are given in Table 2.

The difference between the acceleration mechanisms of Cockcroft and Walton and Ising depend upon whether the fields are static (i.e. conservative) or time-varying (i.e. non-conservative). The electric field can be expressed in a very general form as the sum of two terms, the first being derived from a scalar potential and the second from a vector potential,

$$\vec{E} = -\nabla\phi - \frac{\partial}{\partial t}\vec{A} \quad (1)$$

where

$$\vec{B} = \nabla \times \vec{A} . \quad (2)$$

**Table 2**  
The second "History Line"

1924	Ising proposes time-varying fields across drift tubes. This is "resonant acceleration", which can achieve energies above that given by the highest voltage in the system.
1928	Wideröe demonstrates Ising's principle with a 1 MHz, 25 kV oscillator to make 50 keV potassium ions.
1929	Lawrence, inspired by Wideröe and Ising, conceives the cyclotron.
1931	Livingston demonstrates the cyclotron by accelerating hydrogen ions to 80 keV.
1932	Lawrence's cyclotron produces 1.25 MeV protons and he also splits the atom just a few weeks after Cockcroft and Walton (Lawrence received the Nobel Prize in 1939).

The first term in (1) describes the static electric field of the Cockcroft-Walton and Van de Graaff machines. When a particle travels from one point to another in an electrostatic field, it gains energy according to the potential difference, but if it returns to the original point, for example, by making a full turn in a circular accelerator, it must return to its original potential and will lose exactly the energy it has gained. Thus a gap with a DC voltage has no net accelerating effect in a circular machine.

The second term in (1) describes the time-varying field. This is the term that makes all the present-day high-energy accelerators function. The combination of (1) and (2) yields Faraday's law,

$$\nabla \times \vec{E} = -\frac{\partial}{\partial t}\vec{B} ,$$

which relates the electric field to the rate of change of the magnetic field. There are two basic geometries used to exploit Faraday's Law for acceleration. The first of which is the basis of Ising's idea and the second "history line", and the second is the basis of the third "history line" to be described later.

Ising suggested accelerating particles with a linear series of conducting drift tubes and Wideröe built a 'proof-of-principle' linear accelerator in 1928 [5]. Alternate drift tubes are connected to the same terminal of an RF generator. The generator frequency is adjusted so that a particle traversing a gap sees an electric field in the direction of its motion and while the particle is inside the drift tube the field reverses so that it is again the direction of motion at the next gap. As the particle gains energy and speed the structure periods must be made longer to maintain synchronism (see Fig. 4).

Clearly, as the velocity increases the drift tubes become inconveniently long, unless the frequency can be increased, but at high frequencies the open drift-tube structure is lossy. This problem is overcome by enclosing the structure to form a cavity (in a circular machine) or series

of cavities (in a linear machine), working typically in the MHz range. The underlying principle remains unchanged, but there are several variants of the accelerating structure design.

Ising's original idea can be considered as the beginning of the '**true' accelerator**. Indeed, the next generation of linear colliders, which will be in the TeV range, will probably still be applying his principle of resonant acceleration, except that the frequency will probably be in the tens of GHz range.

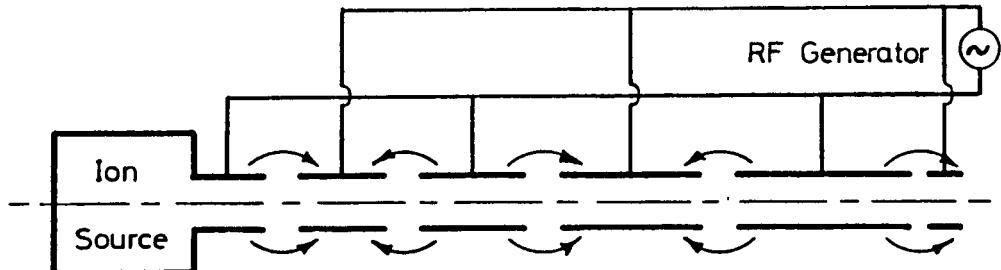


Fig. 4 RF linac

Technologically the linear accelerator, or **linac** as it is known, was rather difficult to build and, during the 1930's, it was pushed into the background by a simpler idea conceived by Ernest Lawrence in 1929 [6], the fixed-frequency cyclotron (see Fig. 5). Lawrence's idea was inspired by a written account of Wideröe's work and M. Livingston demonstrated the principle by accelerating hydrogen ions to 80 keV in 1931. Lawrence's first model worked in 1932 [7]. It was less than a foot in diameter and could accelerate protons to 1.25 MeV. He split the atom only weeks after Cockcroft and Walton. Lawrence received the Nobel Prize in 1939, and by that year the University of California had a 5-foot diameter cyclotron (the 'Crocker' cyclotron) capable of delivering 20 MeV protons, twice the energy of the most energetic alpha particles emitted from radioactive sources. The cyclotron, however, was limited in energy by relativistic effects and despite the development of the synchrocyclotron, a new idea was still required to reach yet higher energies in order to satisfy the curiosity of the particle physicists. This new idea was to be the **synchrotron**, which will be described later.

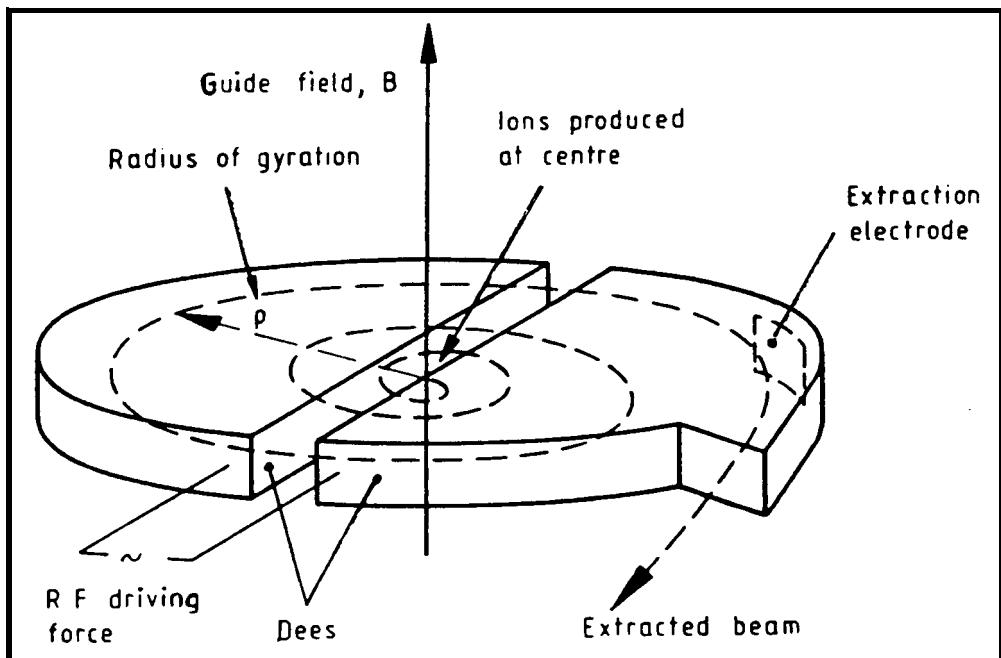


Fig. 5 Schematic cyclotron

### 2.3 The third and fainter 'History Line'

In the previous section, it was mentioned that there were **two** equipment configurations for exploiting Faraday's Law for acceleration. First, consider the application of Faraday's Law to the linac, which is made more evident by enclosing the gaps in cavities. For simplicity the fields in a single RF cavity are shown schematically in Fig. 6(a).

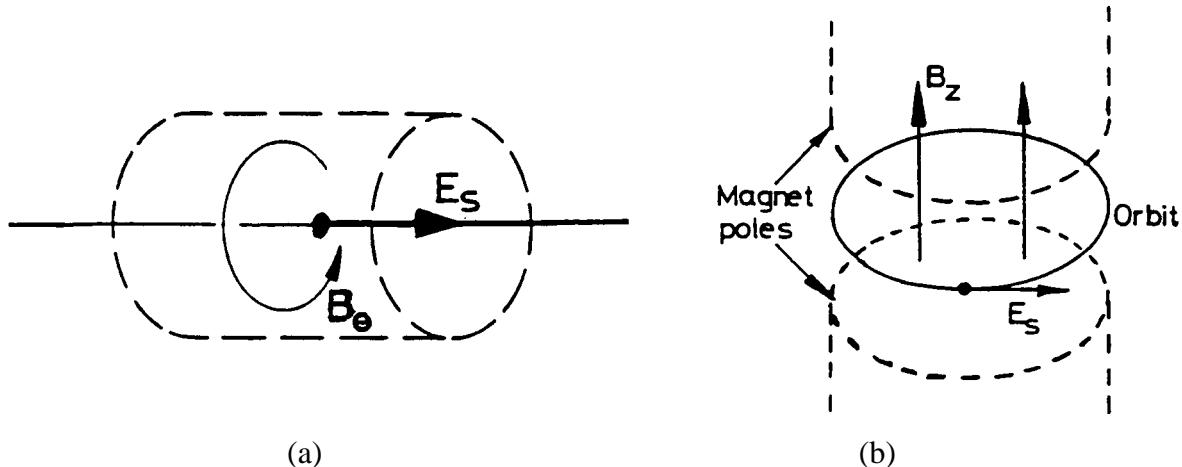


Fig. 6 Acceleration configurations

The azimuthal magnetic field is concentrated towards the outer wall and links the beam. Faraday's Law tells us the periodic rise and fall of this magnetic field induces an electric field on the cavity axis, which can be synchronised with the passage of the beam pulse.

Suppose now that the topology is transformed, so that the beam encircles the magnetic field as shown in Fig. 6(b). Wideröe [8, 9] suggested this configuration and the acceleration mechanism, now known as "betatron acceleration". He called his idea a "strahlung transformator" or "ray transformer", because the beam effectively formed the secondary winding of a transformer (see Figs. 6 and 7)). As the flux through the magnet core is increased, it induces an azimuthal e.m.f. which drives the charged beam particles to higher and higher energies. The trick is to arrange for the increase in the magnetic field in the vicinity of the beam to correspond to the increase in particle energy, so that the beam stays on the same orbit\*. This device, the betatron, is insensitive to relativistic effects and was therefore ideal for accelerating electrons. The betatron has also the great advantages of being robust and simple. The one active element is the power converter that drives the large inductive load of the main magnet. The focusing and synchronisation of the beam energy with the field level are both determined by the geometry of the main magnet. As noted in the third "history line" in Table 3, Wideröe put this idea in his laboratory notebook, while he was a student, but it remained unpublished only to re-surface many years later when Kerst [10] built the first machine of this type. When in 1941 Kerst and Serber published a paper on the particle oscillation in their betatron [11], the term "betatron oscillation" became universally adopted for referring to such oscillations in all devices.

---

\* Known as the Wideröe condition, or 2-to-1 rule.

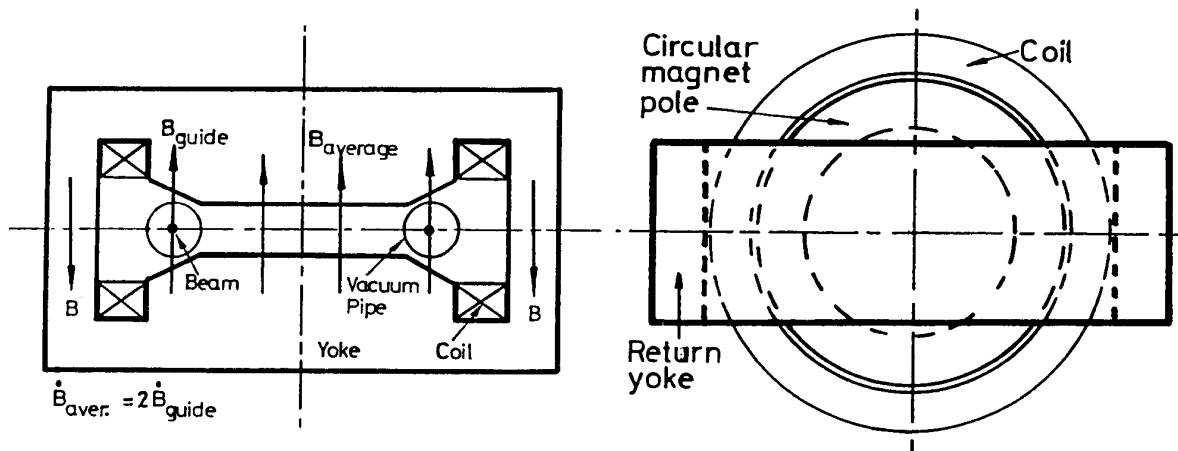


Fig. 7 Strahlung transformator or betatron

**Table 3**  
The third "History Line"

1923	Wideröe, a young Norwegian student, draws in his laboratory notebook the design of the betatron with the well-known 2-to-1 rule. Two years later he adds the condition for radial stability <b>but does not publish.</b>
1927	Later in Aachen Wideröe makes a model betatron, but it does not work. Discouraged he changes course and builds the linear accelerator mentioned in Table 2.
1940	Kerst re-invents the betatron and builds the first working machine for 2.2 MeV electrons.
1950	Kerst builds the world's largest betatron of 300 MeV.

The development of betatrons for high-energy physics was short, ending in 1950 when Kerst built the world's largest betatron (300 MeV), but they continued to be built commercially for hospitals and small laboratories where they were considered as reliable and cheap. In fact the betatron acceleration mechanism is still of prime importance. In the present-day synchrotron, there is a small contribution to the beam's acceleration which arises from the increasing field in the main dipoles. If an accurate description of the longitudinal motion is required, then the betatron effect has to be included.

### 3. THE MAIN DEVELOPMENT

By the 1940's three acceleration mechanisms had been demonstrated:- **DC acceleration, resonant acceleration and the betatron mechanism.** In fact, there were to be no new ideas for acceleration mechanisms until the mid-1960's, when collective acceleration [12] was proposed in which heavy ions are accelerated in the potential well of an electron ring and the 1980's when there were several Workshops devoted entirely to finding new acceleration techniques. However, the acceleration mechanism is not sufficient by itself and other equally important developments are needed.

In order to accelerate particles to very high energies, it is also necessary to have focusing mechanisms in the transverse and longitudinal (energy) planes. This was not always appreciated. In the early cyclotrons, for example, the field was made as uniform as possible only to find that the beam was unstable. Livingston [13] who was Lawrence's research student, told how they shimmed the magnet for each small step in energy to keep the beam stable, thus ending up with a field shape for transverse stability that decreased with radius. Theory has later shown that this decrease should be an inverse power law of the radius between zero and unity.

The cyclotron is limited by relativistic effects, which cause the particles to slow down and lose synchronism with the RF field. At first glance it would appear that one would only have to reduce the frequency in order to maintain synchronism, but this is a little too naïve since the spread in revolution frequency with energy would quickly exploit the natural energy spread in the beam and disperse the particles away from the peak of the RF voltage. In this case a longitudinal focusing mechanism is needed. This problem was overcome by E. McMillan [14] and independently by V. Veksler [15] who discovered the principle of **phase stability** in 1944 and invented the **synchrotron**. Phase stability is general to all RF accelerators except the fixed-frequency cyclotron. The effect is that a bunch of particles, with an energy spread, can be kept bunched throughout the acceleration cycle by simply injecting them at a suitable phase of the RF cycle. This focusing effect was strong enough that the frequency modulation in the synchro-cyclotron did not have to be specially tailored and was simply sinusoidal. Synchro-cyclotrons can accelerate protons to about 1 GeV, a great improvement on the simple cyclotron, but the repetition rate reduces the particle yield.

In the synchrotron [14, 15] the guide field increases with particle energy, so as to keep the orbit stationary as in the betatron, but acceleration is applied with an RF voltage via a gap or cavity. In 1946 F. Goward and D. Barnes [16] were the first to make a synchrotron work, and in 1947 M. Oliphant, J. Gooden and G. Hyde [17] proposed the first proton synchrotron for 1 GeV in Birmingham, UK. However, the Brookhaven National Laboratory, USA, built their 3 GeV Cosmotron by 1952, just one year ahead of the Birmingham group.

Up to this time the only mechanism known for focusing in the transverse plane was called **weak, or constant-gradient focusing**. In this case, the guide field decreases slightly with increasing radius and its gradient is constant all round the circumference of the machine. The tolerance on the gradient is severe and sets a limit to the size of such an accelerator. The aperture needed to contain the beam also becomes very large and the magnet correspondingly bulky and costly. In the early fifties the limit was believed to be around 10 GeV.

In the same year as the Cosmotron was finished (1952) E. Courant, M. Livingston and H. Snyder [18] proposed **strong focusing**, also known as **alternating-gradient (AG) focusing**. The idea had been suggested earlier by Christofilos [19] but it was not published. This new principle revolutionized synchrotron design, allowing smaller magnets to be used and higher energies to be envisaged. It is directly analogous to a well-known result in geometrical optics, that the combined focal length  $F$  of a pair of lenses of focal lengths  $f_1$  and  $f_2$  separated by a distance  $d$  is given by

$$\frac{1}{F} = \frac{1}{f_1} + \frac{1}{f_2} - \frac{d}{f_1 f_2}.$$

If the lenses have equal and opposite focal lengths,  $f_1 = -f_2$  and the overall focal length  $F = f^2/d$ , which is always positive. In fact,  $F$  remains positive over quite a large range of values when  $f_1$  and  $f_2$  have unequal values but are still of opposite sign. Thus within certain limits a series of alternating lenses will focus. Intuitively one sees that, although the beam may be defocused by one lens, it arrives at the following lens further from the axis and is therefore focused more strongly. Structures based on this principle are referred to as AG structures.

The synchrotron quickly overshadowed the synchrocyclotron and the betatron in the race for higher energies. The adoption of alternating gradient focusing for machines and transfer lines was even quicker. CERN for example immediately abandoned its already-approved project for a 10 GeV/c weak focusing synchrotron in favour of a 25 GeV/c AG machine, which it estimated could be built for the same price.

The next step was the **storage ring collider**. In physics experiments, the useful energy for new particle production is the energy that is liberated in the centre-of-mass system. When an accelerator beam is used on a fixed target, only a fraction of the particle's energy appears in the centre-of-mass system, whereas for two equal particles in a head-on collision, all of the particles' energy is available. This fundamental drawback of the fixed-target accelerator becomes more punitive as the energy increases. For example, it would have needed a fixed-target accelerator of over 1TeV to match the centre-of-mass energy available in the CERN ISR (2 x 26 GeV proton colliding rings).

The storage-ring collider now dominates the high-energy physics field. Single-ring colliders, using particles and antiparticles in the same magnetic channel, were the first type of collider to be exploited at Frascati in the AdA (Annelli di Accumulazione) project (1961). The first double-ring proton collider was the CERN ISR (Intersecting Storage Rings), 1972-1983. The highest-energy collisions obtained to date are 2 x 900 GeV in the Fermilab, single-ring, proton-antiproton collider.

Colliders have been very successful as physics research instruments. The  $J/\psi$  particle was discovered at SPEAR by B. Richter and at the same time by Ting at BNL – they shared the 1976 Nobel Prize. The CERN proton-antiproton storage ring was also the source of a Nobel Prize for C. Rubbia and S. van der Meer in 1984, following the discovery of the W and Z particles. The proton-antiproton colliders were only made possible by the invention of **stochastic cooling** by S. van der Meer for the accumulation of the antiprotons [20].

The use of **superconductivity** in proton machines has made the very highest energies possible. There has also been another change taking place, which has been called the **Exo-geographical transition** (a phrase coined by Professor N. Cabibbo at a Workshop held at Frascati in 1984). This refers to the arrangements that have made it possible to bury the very large machines such as LEP and HERA deep under property which does not belong to the laboratory concerned. Without such agreements, Europe could not have maintained its leading position in the world accelerator league.

In order to fill in some of the bigger gaps in this brief history, it is now necessary to jump back in time to mention some of the other accelerators, which may not have featured as a high-energy machine, but have found their place as injectors or as being suitable for some special application.

The **microtron**, sometimes known as the electron cyclotron, was an ingenious idea due to Veksler (1945). The electrons follow circular orbits of increasing radius, but with a common tangent. An RF cavity positioned at the point of the common tangent supplies a constant energy increment on each passage. The relativistic mass increase slows the revolution frequency of the electrons, but by a constant increment on each passage. If this increment is a multiple of the RF oscillator frequency, the electrons stay in phase, but on a different orbit. Microtrons operate at microwave frequencies and are limited to tens of MeV. They are available commercially and are sometimes used as an injector to a larger machine.

The **radio-frequency quadrupole** (RFQ) suggested in 1970 by I. Kapchinski and V. Telyakov is useful at low energies and is increasingly replacing the Cockcroft-Walton as injector. The RFQ combines focusing and acceleration in the same RF field.

The electron storage rings have given birth to the synchrotron radiation sources, more usually referred to as **light sources**. These machines are now the fastest growing community in the accelerator world and the first commercially available compact synchrotron light source for lithography has just come onto the market.

The linear accelerator was eclipsed during the thirties by circular machines. However, the advances in ultra-high frequency technology during World War II (radar) opened up new possibilities and renewed interest in linac structures. Berkeley was first, with a proton linear accelerator of 32 MeV built by Alvarez in 1946. The **Alvarez accelerator** has become very popular as an injector for large proton and heavy-ion synchrotrons all over the world with energies in the range of 50–200 MeV, that is essentially non-relativistic particles. The largest proton linear accelerator to date is the 800 MeV 'pion factory' (LAMPF) at Los Alamos.

The first electron linear accelerators were studied at Stanford and at the Massachusetts Institute for Technology (MIT) in 1946. This type of accelerator has also had a spectacular development, up to the largest now in operation, the 50 GeV linear accelerator at the Stanford Linear Accelerator Centre (SLAC). Like betatrons they have become very popular in fields outside nuclear physics, particularly for medicine.

The Livingston chart (see Fig. 8) shows, in a very striking way, how the succession of new ideas and new technologies has relentlessly pushed up accelerator beam energies over five decades at the rate of over one and a half orders of magnitude per decade. One repeatedly sees a new idea, which rapidly increases the available beam energy, but only to be surpassed by yet another new idea. Meanwhile the first idea continues into saturation and possibly into quasi-oblivion.

This brings the section on the main development almost up to date, except for the Stanford Linear Collider (SLC), but this will be mentioned under future accelerators where it fits more naturally.

#### 4. THE CURRENT SITUATION IN HIGH-ENERGY PARTICLE PHYSICS ACCELERATORS

Table 4 contains a selection of the main operating high-energy physics machines, those under construction and those under study. The latter two groups encompass the extremes of machines like RHIC [22], which are partially constructed and the linear colliders, which are very futuristic.

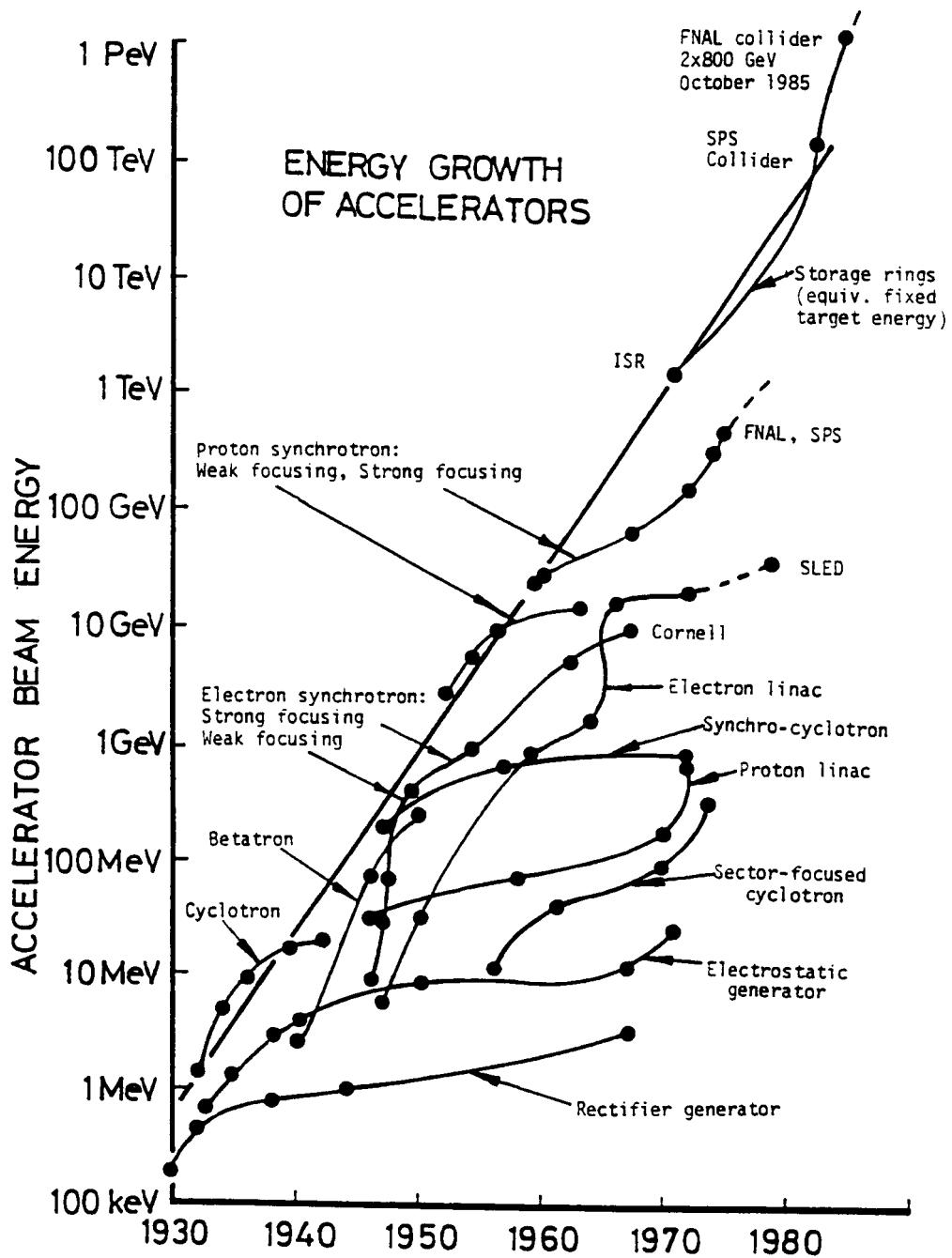


Fig. 8 Livingston chart [21]

In the present situation circular colliders dominate the high-energy field. The proton-antiproton colliders are now mature machines and it is unlikely that the USA or Western Europe will propose further facilities of this type. The technologies of stochastic and electron cooling that were developed for this class of facility are now being applied in smaller storage rings.

Once LEP [23] has been upgraded to around 100 GeV, it will almost certainly be the highest energy electron ring to be built, since the penalty to be paid in RF power to compensate the synchrotron radiation loss is already prohibitive at this energy. The solution is to change to **linear electron colliders**; a solution that was already foreseen in 1965 by Tigner [24]. The Stanford Linear Collider (SLC) [25] is a test bed for these future machines.

**Table 4**  
Operating high-energy physics accelerators

Accelerator	Particles	Beam energy [GeV]	c.m. energy [GeV])	Luminosity [cm <sup>-2</sup> s <sup>-1</sup> ]	Remarks
KEK Japan	p	12	5	-	Fixed target
AGS Brookhaven	p	33	8	-	Fixed target Polarised p
PS CERN	p e <sup>+</sup> , e <sup>-</sup> , p <sup>-</sup> , ions	28 (p) 3.5 (e)	7 -	-	Fixed target Injector
CESR Cornell	e <sup>+</sup> , e <sup>-</sup>	9	18	10 <sup>32</sup>	Collider
Tevatron II FNAL	p p, p <sup>-</sup>	800 (p)	40 -	-	Fixed target Injector
SPS CERN	p, e p, p <sup>-</sup>	450 (p), 20 (e) 2x315	30 (p),- 630	- 3 x 10 <sup>30</sup>	F. target, injector Collider
SLC SLAC	e <sup>+</sup> , e <sup>-</sup>		100	(6 x 10 <sup>30</sup> )	Linear collider
Tevatron I FNAL	p, p <sup>-</sup>	900	1800	10 <sup>31</sup>	s.c. collider
TRISTAN in Japan	e <sup>+</sup> , e <sup>-</sup>	32	64	8 x 10 <sup>31</sup>	Collider s.c. cavities
LEPI CERN	e <sup>+</sup> , e <sup>-</sup>	55	110	1.6 x 10 <sup>31</sup>	Collider
HERA DESY	e, p	30 (e <sup>-</sup> ) 820 (p)	310	3 x 10 <sup>31</sup>	Collider s.c. p-ring

High-energy physics accelerators under construction

UNK I USSR	p	400	28	-	Fixed target Conventional
SSC USA	p, p	20	40	~ 10 <sup>33</sup>	s.c. collider
LEP II	e <sup>+</sup> , e <sup>-</sup>	100	200	10 <sup>32</sup>	Collider, s.c. cavity upgrade
RHIC Brookhaven	p to Au	0.25 to 0.1/amu	0.5 to 0.2/amu	3 x 10 <sup>30</sup> 1.2 x 10 <sup>27</sup>	s.c. collider for heavy ions

High-energy physics accelerators under study

UNK II USSR	p, p p, p <sup>-</sup>	3	6	~ 4 x 10 <sup>32</sup> ~ 10 <sup>37</sup>	s.c. collider for 1996
LHC CERN	p, p	8	16	~ 10 <sup>34</sup>	s.c. collider
CLIC CERN	e <sup>+</sup> , e <sup>-</sup>	1	2	~ 10 <sup>33</sup>	Linear collider
SC Stanford	e <sup>+</sup> , e <sup>-</sup>	0.5 (1)	1(2)	~ 10 <sup>33</sup>	Linear collider proposal 1990
VLEPP USSR	e <sup>+</sup> , e <sup>-</sup>	0.5 (1)	1 (2)	~ 10 <sup>33</sup>	Linear collider for 1996
JLC Japan	e <sup>+</sup> , e <sup>-</sup>	0.5	1	~ 10 <sup>33</sup>	-

At the moment, the proton community is poised to build the SSC (Super Superconducting Collider) in Texas [26] and the LHC (Large Hadron Collider) in CERN [27]. Both machines are superconducting and of very large dimensions. At present there is no hard limitation on the size of hadron colliders, except of course cost. However, synchrotron radiation is already a bothersome heat load in these machines and will be a very real problem in machines of the size of the Eloisatron [28] for example. The LHC is a high technology project, which will use high-field magnets (approaching 10 T) with probably 'niobium-titanium' technology at 2 K in the arcs and niobium-tin technology at 4 K in the insertions. The magnets will also be of the twin-bore design first proposed by Blewett [29].

## 5. CONCLUSION

Led by the example of the SLC, accelerator builders are now tackling formidable theoretical and technological problems in all stages of the accelerator design. In the next generation of proposed linear electron colliders the typical values required for the normalised emittance are of the order of  $10^{-7}$  rad.m. Beam sizes at the interaction point will have to be around 1 to 30 nm high with pulse lengths of 200-800  $\mu$ m. With a few  $10^{10}$  particles per bunch and a repetition rate of over 1000 Hz the beam power is then a few MW. The luminosity in such designs is around a few times  $10^{33}$  cm $^{-2}$  s $^{-1}$ , far higher than anything that has yet been achieved. Stability of the supporting structures and power converters driving the final focus become critically important with such small beam sizes and the fabrication of elements such as the final focus quadrupoles requires new techniques.

At present the linear collider designs are called quasi-conventional. For example, the CERN CLIC study [30] assumes the use of a warm copper accelerating structure operating at 29 GHz giving 80 MV/m. If this sounds easy, then consider that the structure will be powered from a superconducting drive linac. Such high-gradient, high-frequency structures have never before been used and neither has a superconducting linac been used in this way to drive a second accelerator. In fact, the term "quasi-conventional" is really a misnomer.

The future holds many challenges for the accelerator engineer both in the gigantic superconducting hadron machines now proposed and in the new generation of electron linear colliders.

\* \* \*

## REFERENCES

- [1] R.W. Gurney and E.U. Condon, *Nature* 122, 439 (1928) and  
G. Gamov, *Zeit f. Phys.* 51, 204 (1928).
- [2] J.D. Crockcroft and E.T.S. Walton, "Experiments with high velocity ions", *Proc. Royal Soc., Series A* 136 (1932), 619–30.
- [3] R.J. Van de Graaff, "A 1,500,000 volt electrostatic generator", *Phys. Rev.*, 387, (Nov. 1931), 1919–20.
- [4] G. Ising, *Arkiv för Matematik, Astronomi och Fysik*, 18 (1924), 1–4.
- [5] R. Wideröe, *Arch.für Elektrotechnik*, 21 (1928), 387–406.

- [6] E.O. Lawrence and N.E. Edlefsen, *Science*, 72 (1930), 376–7.
- [7] E.O. Lawrence and M.S. Livingston, "The production of high speed light ions without the use of high voltages", *Phys. Rev.*, 40 (April 1932), 19–35.
- [8] W. Paul, "Early days in the development of accelerators", *Proc. Int. Symposium in Honour of Robert R. Wilson*, Fermilab, 1979 (Sleepeck Printing Co. Bellwood, Ill., 1979), 25–688.
- [9] R. Wideröe, "Some memories and dreams from the childhood of particle accelerators", *Europysics News*, Vol. 15, No. 2 (Feb. 1984), 9–11.
- [10] D.W. Kerst, "The acceleration of electrons by magnetic induction", *Phys. Rev.*, 60 (July 1942), 47–53.
- [11] D.W. Kerst and R. Serber, "Electronic orbits in the induction accelerator", *Phys. Rev.*, 60 (July 1941), 53–58.
- [12] G.S. James, R.H. Levy, H.A. Bethe and B.T. Fields, "On a new type of accelerator for heavy ions", *Phys. Rev.*, 145, 925 (1966).
- [13] M.S. Livingston, "Particle accelerators: A brief history", Harvard University Press, Cambridge, Massachusetts (1969), 32.
- [14] E.M. McMillan, "The synchrotron - a proposed high-energy particle accelerator", *Phys. Rev.*, Letter to the editor, 68 (Sept. 1945), 1434.
- [15] V. Veksler, *J. of Phys. USSR*, 9 (1945), 153–8.
- [16] F.K. Goward and D.E. Barnes, *Nature*, 158 (1946), 413.
- [17] M.L. Oliphant, J.S. Gooden and G.S. Hyde, *Proc. Phys. Soc.*, 59 (1947), 666.
- [18] E.D. Courant, M.S. Livingston and H.S. Snyder, "The strong-focusing synchrotron – a new high-energy accelerator", *Phys. Rev.*, 88 (Dec. 1952), 1190–6, and E.D. Courant and H.S. Snyder, "Theory of the alternating-gradient synchrotron", *Annals of Physics*, No. 3 (1958), 1–48.
- [19] N.C. Christofilos, Unpublished report (1950), and U.S. Patent no. 2.736,799, filed March 10, 1950, issued February 28, 1956.
- [20] S. van der Meer, "Stochastic damping of betatron oscillations in the ISR", CERN/ISR-PO/72-31 (August, 1972).
- [21] Livingston chart. First published in Livingston's book: "High-energy accelerators", Interscience Publishers Inc., New York (1954).
- [22] H. Hahn, "The relativistic heavy ion collider at Brookhaven", *Proc. 1st European Part. Accel. Conf. (EPAC)*, Rome 1988 (World Scientific, Singapore, 1989), 109–11.
- [23] E. Picasso, "The LEP project", *Proc. 1st European Part. Accel. Conf. (EPAC)*, Rome 1988 (World Scientific, Singapore, 1989) 3–6.
- [24] M. Tigner, "A possible apparatus for electron clashing-beam experiments", Letter to the editor, *Nuovo Cim.*, 37 (1965), 1228–31.

- [25] R. Stiening, "The status of the Stanford Linear Collider", Proc. 1987 IEEE Part. Accel. Conf., Washington, March 1987, (IEEE, 1987), 1–7.
- [26] J. Peoples, "The SSC Project", *Ibid* 237–41.
- [27] G. Brianti, "The large hadron collider (LHC) in the LEP tunnel", *Ibid*. 218–222.
- [28] K. Johnsen, "Long term future", ECFA Study Week on Instr. Tech. for High Luminosity Hadron Colliders, Barcelona, 1989, CERN 89–10, ECFA 89–124, Vol. 1 (Nov. 1989), 25–34.
- [29] J.P. Blewett, "200 TeV Intersecting Storage Accelerators", 8th Int. Conf. on High-Energy Accel., CERN, 1971 (CERN, Geneva, 1971), 501–4.
- [30] W. Schnell, "Linear Collider Studies", Proc. 1st European Part. Accel. Conf. (EPAC), Rome 1988 (World Scientific, Singapore, 1989), 285–89.

# BASIC COURSE ON ACCELERATOR OPTICS

J. Rossbach, Deutsches Elektronen-Synchrotron DESY

P. Schmüser, II. Institut für Experimentalphysik, Universität Hamburg, F.R. Germany

## Abstract

The main subject of this introductory course is transverse optics in cyclic high energy accelerators. It is based on three one-hour lectures on “Basic transverse optics” given at the CERN Accelerator Schools (P. Schmüser, 1986 [5], 1988, 1990, and J. Rossbach, 1992). Although the emphasis is on periodic solutions and their stability, application of the formalism to non-periodic structures is also treated. We have not expanded the material beyond what can be covered in 3 lectures, but the proofs are done more carefully and some examples and details have been added which might be useful in practice.

After a brief introduction to the concepts of both weak and strong focusing of relativistic particle beams, types of magnetic lenses are discussed, followed by a careful treatment of the multipole expansion of transverse magnetic fields. Linear transverse single-particle dynamics is treated both in terms of transfer matrices and betatron oscillations. Transfer matrices of the most common accelerator lattice modules are explicitly given, including a brief introduction to quadrupole doublet imaging. Dispersion is treated, but no linear coupling. The concept of beam emittance, including Liouville’s Theorem, is discussed from different points of view. Also effects of linear field errors, stability criteria and chromaticity are dealt with. FODO cells are treated in various respects. Finally, strong and weak focusing are compared quantitatively.

## 1 INTRODUCTION

In any kind of accelerator there is exactly one curve – the design orbit – on which ideally all particles should move, see Fig. 1. If this design orbit is curved, which may be required for many reasons, *bending* forces are needed.

In reality, most particles of the beam will deviate slightly from the design orbit. In order to keep these deviations small on the whole way (which might be as long as  $10^{11}$  km in a storage ring), *focusing* forces are required. Both bending and focusing forces can be accomplished with electromagnetic fields. The Lorentz force is

$$\mathbf{F} = e(\mathbf{E} + \mathbf{v} \times \mathbf{B})$$

For velocities  $v \approx c$ , a moderate magnetic field of 1 Tesla corresponds to a huge electric field of  $3 \cdot 10^8$  V/m. Therefore only transverse magnetic fields are considered.

Now consider the total energy  $E$  of a particle:

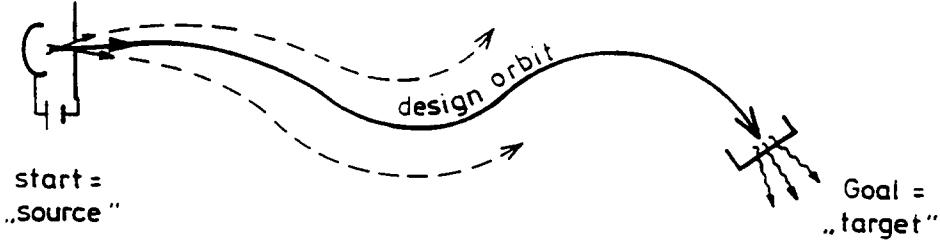


Figure 1: To guide particles on a curved design orbit, bending elements are required.

$$E = \frac{m_0 c^2}{\sqrt{1 - \frac{v^2}{c^2}}}$$

Relativistic mechanics describes its rate of change in the presence of an external force by

$$\frac{dE}{dt} = \mathbf{v} \cdot \mathbf{F}$$

Since  $\mathbf{v}$  is perpendicular to  $\mathbf{v} \times \mathbf{B}$ , the energy and the relativistic mass  $m = m_0 / \sqrt{1 - v^2/c^2}$  are not changed in static magnetic fields. So the equation of motion is

$$\dot{\mathbf{v}} = \frac{e}{m} \mathbf{v} \times \mathbf{B} \quad (1.1)$$

The particles have to make a large number of revolutions in a circular accelerator or storage ring. Stability of motion is an important criterion and puts stringent requirements on the magnetic field in the vicinity of the equilibrium orbit. Depending on the magnitude of the focusing forces we can distinguish between weak and strong focusing.

The first machines were built with weak focusing and a guide field which does not depend on the azimuthal angle. If  $\mathbf{B}$  is independent of the azimuth angle  $\Theta$  the equilibrium orbit is a circle of radius  $\rho = mv/eB$  which we choose to be in the plane  $z = 0$ . The motion is stable if for small deviations of the particles from this orbit restoring forces arise which lead to oscillations around the orbit. These are called betatron oscillations.

Consider the motion in the horizontal plane  $z = 0$  containing the equilibrium orbit (Fig. 2). Stability requires that the Lorentz force is smaller than the centrifugal force for  $r < \rho$  and larger than it for  $r > \rho$

$$evB_z(r) \begin{cases} < \frac{mv^2}{r} & \text{for } r < \rho \\ > \frac{mv^2}{r} & \text{for } r > \rho \end{cases} \quad (1.2)$$

For small deviations from the equilibrium orbit we have

$$r = \rho + x = \rho \left(1 + \frac{x}{\rho}\right) \quad (1.3)$$

$$\frac{mv^2}{r} \approx \frac{mv^2}{\rho} \left(1 - \frac{x}{\rho}\right)$$

$$evB_z(r) \approx ecB_0 \left(1 - n \frac{x}{\rho}\right) \quad (1.4)$$

with  $B_0 = B_z(\rho)$ . In Equation (1.4) we have introduced the *field index*  $n$

$$n = -\frac{\rho}{B_0} \left( \frac{\partial B_z}{\partial r} \right)_{r=\rho} \quad (1.5)$$

Inserting (1.3) and (1.4) into (1.2) we obtain the condition  $n < 1$ .

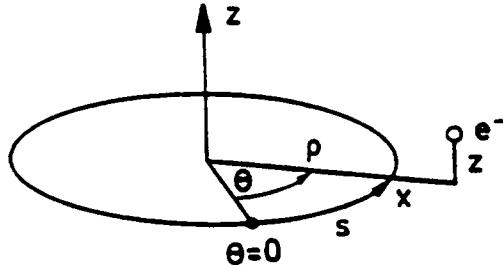


Figure 2: Circular design orbit and coordinates used to describe particle trajectories

It is worth noting that Eq. (1.2) is satisfied even with  $n \equiv 0$ , that is with a homogeneous field. This reflects the fact, that in a homogeneous field all plane orbits are circles, so that particles diverging from one point will meet again after  $180^\circ$  of revolution (see Fig. 3).

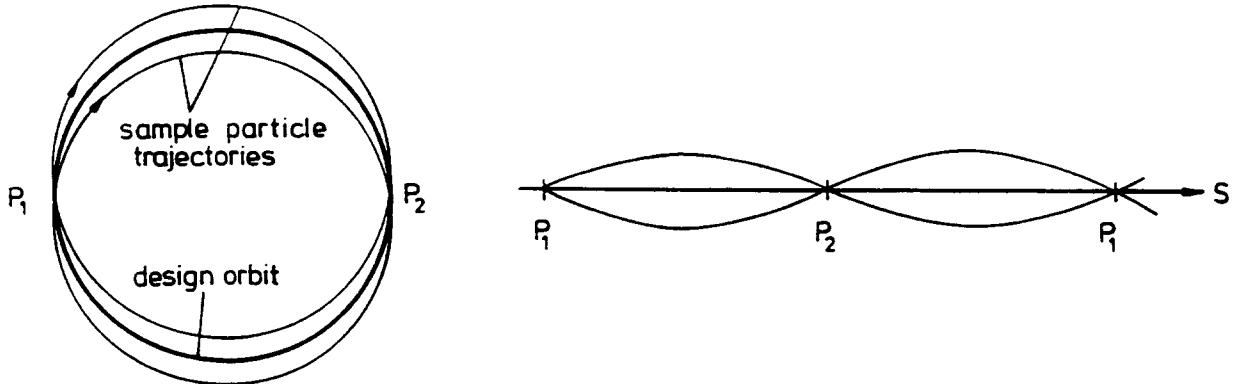


Figure 3: Geometrical focusing in a homogeneous magnetic field. All particles starting at  $P_1$  with slightly different angles  $\alpha$  move on circles with the same radii and meet again at  $P_2$ . With respect to the design orbit, this looks like focusing, which can be seen more clearly from the right-hand part of the figure where the design orbit has been developed into a straight line. The maximum deviation from the design orbit (i.e. the amplitude of betatron oscillations) is easily estimated at  $\alpha \cdot \rho$ . Thus for a typical beam divergence of  $\alpha = 1 \text{ mrad}$  it will be 1 mm at  $\rho = 1 \text{ m}$  but as large as 1000 mm in a high energy ring with  $\rho = 1000 \text{ m}$ .

This is called geometrical focusing, and it is an essential contribution to the weak focusing in the horizontal plane. It is, however, not sufficient:

Consider a particle in a homogeneous field with initial conditions such that it moves on a circle in the horizontal plane  $z \equiv 0$ . If it now gets an infinitely small kick into the vertical direction, it will spiral away to infinity and will be lost. To achieve stability in the vertical plane we need a restoring force

$$F_z = -C \cdot z \quad (C = \text{const.})$$

and therefore a horizontal field component

$$B_x = -C' \cdot z$$

Now from  $\nabla \times \mathbf{B} = 0$ :  $\partial B_x / \partial z = \partial B_z / \partial x = \partial B_z / \partial r$ . So  $B_z$  has to decrease with increasing  $r$ , see Fig. 4.  $\partial B_z / \partial r < 0$  is equivalent to a positive field index,  $n > 0$ .

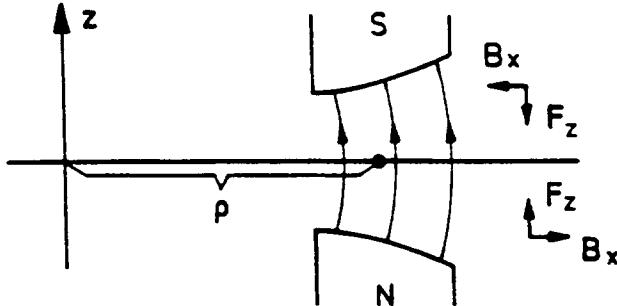


Figure 4: Shape of magnetic field around design orbit

Stability in both horizontal and vertical motion is achieved if the field index fulfills the inequalities

$$0 < n < 1 \quad (1.6)$$

The restoring forces are linear in the deviations  $x$  and  $z$  from the equilibrium orbit, if  $x$  and  $z$  are small. We therefore obtain harmonic oscillations. The differential equations will not be derived here, because they are special cases of the equations in section 3.

$$\begin{aligned} \ddot{x} + \omega_0^2(1-n)x &= 0 \\ \ddot{z} + \omega_0^2 n \cdot z &= 0 \end{aligned} \quad (1.7)$$

with  $\omega_0 = eB_0/m$ .  $f_0 = \omega_0/2\pi$  is the revolution frequency (cyclotron frequency) of the particles. Since  $0 < n < 1$  we see that in a weakly-focusing machine the frequency of betatron oscillations is lower than the revolution frequency, i.e. there is less than one oscillation per revolution.

$$f_x = \sqrt{1-n} f_0, \quad f_z = \sqrt{n} f_0 \quad (1.8)$$

The principle of weak focusing has one serious drawback: Since the betatron oscillation wavelength is larger than the circumference of the machine one gets large deviations from the orbit if the circumference is large. The magnet apertures must be very big. The apertures can be drastically reduced if one applies strong focusing ( $|n|$  much larger than 1). This is impossible in a machine which has a guide and focusing field independent of the azimuthal angle, since in that case the condition  $0 < n < 1$  has to hold, as we have just shown. It is, however, possible if we split up the machine into a series of magnetic sectors in which in alternating order the magnetic field increases strongly with increasing radius ( $n \ll -1$ ) or decreases strongly with increasing radius ( $n \gg +1$ ) (see Fig. 5).

An alternating series of focusing and defocusing lenses leads to an overall focusing because the focusing lenses are, on the average, traversed at larger distance from the axis than the defocusing ones. This is shown in Fig. 6. In section 3 we shall see that focusing is indeed obtained in both planes.

The Brookhaven AGS and the CERN PS were the first large alternating-gradient synchrotrons. In the PS the field index is  $n=288$  and there are 6.2 betatron oscillations per revolution. Note

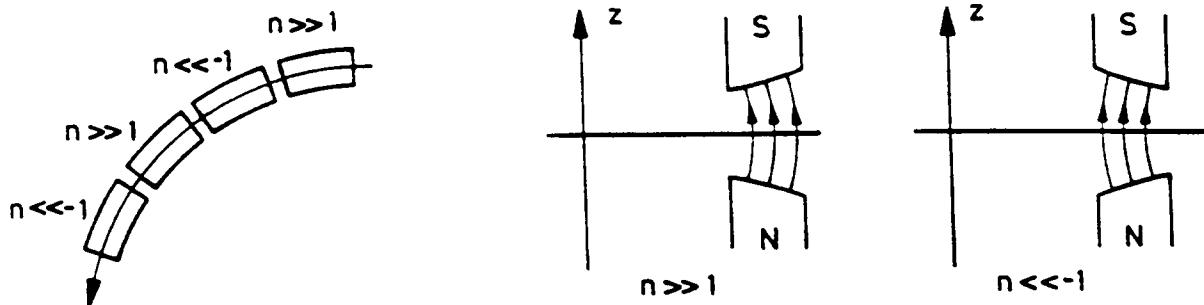


Figure 5: Alternating-gradient focusing

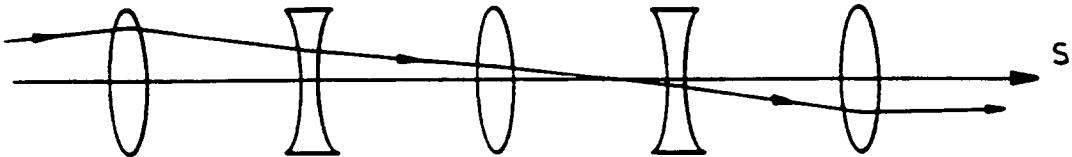


Figure 6: An alternating series of focusing and defocusing lenses leads to overall focusing if the distances between the lenses are not too large.

that Eq. (1.8) is not applicable to strong focusing.

## 2 ACCELERATOR MAGNETS

Many of the older alternating gradient synchrotrons like the CERN proton synchrotron PS or the DESY electron synchrotron have been built with “combined-function” magnets, i.e. magnets which combine a dipole field for deflection and a quadrupole field for focusing. The new large accelerators and storage rings are equipped with “separated-function” magnets: dipoles for deflection, quadrupoles for focusing. These allow higher particle energies because the iron yoke of a pure dipole magnet saturates at higher field levels than the yoke of a combined-function “synchrotron” magnet and moreover there is more flexibility in optimizing the optical properties of the machine.<sup>1</sup>

Traditionally, soft iron of high permeability is used to concentrate the field into the small region where it is needed. This also reduces electric power consumption and easily allows to guarantee high field quality. In those cases where the required field strength is either very small ( $B \ll 0.1$  T) or above the saturation level ( $B > 2$  T), “air coil” magnets are used. Some remarks relevant

<sup>1</sup>In circular electron accelerators, combined-function magnets have another big disadvantage as compared to separated-function magnets. The electrons suffer a permanent energy loss because of the emission of synchrotron radiation which is compensated by the energy gain in the accelerating cavities. In a separated-function electron machine both effects together lead to a damping of the horizontal and vertical betatron oscillations. In a pure combined-function machine, however, only the vertical betatron oscillation is damped whereas the horizontal oscillation will blow up (see the lectures by R. Walker, these proceedings).

for the construction of these magnets are found at the very end of this section.

### a) Dipole magnet

A magnet with flat pole shoes generates a homogeneous field  $B_0$  (Fig. 7).

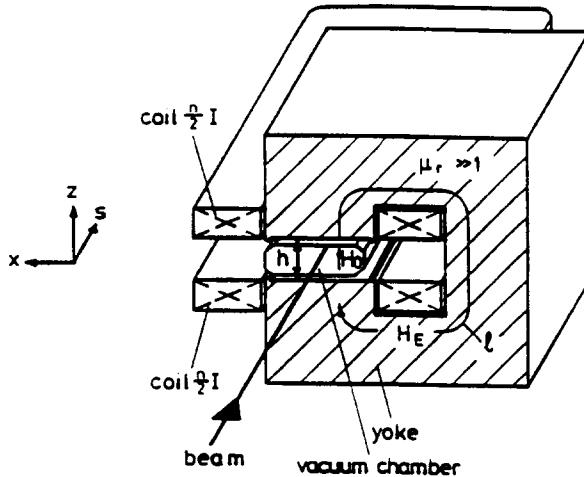


Figure 7: Schematic view of a dipole magnet, showing the path of integration to compute the field in the gap.

The field is computed from the formula

$$\oint \mathbf{H} \cdot d\mathbf{s} = hH_0 + \ell H_E = nI$$

$$H_E = \frac{1}{\mu_r} \cdot H_0$$

For  $\mu_r \gg 1$  we obtain

$$B_0 = \frac{\mu_0 n I}{h} \quad h = \text{gap height} \quad (2.1)$$

Formula (2.1) is only approximate. In particular it neglects fringe fields and iron saturation. The radius of curvature for a particle of charge  $e$  and momentum  $p$  is given by

$$\frac{1}{\rho} [\text{m}^{-1}] = \frac{eB_0}{p} = 0.2998 \frac{B_0 [\text{T}]}{p [\text{GeV}/c]} \quad (2.2)$$

### b) Solenoid lens

A relatively simple magnetic lens arises from the magnetic field of a rotationally symmetric coil, see Fig. 8.

Due to the Maxwell equation  $\operatorname{div} \mathbf{B} = 0$ , the magnetic field, which is purely longitudinal in the inner part of the coil, must contain radial components in the outer part. While particles moving exactly on the axis do not experience any force, the others suffer an azimuthal acceleration due to the radial component while entering and leaving the lens. Because of the azimuthal motion there is a radial force in the longitudinal field. As required for imaging, this force is, indeed, proportional to the radial distance  $r$  if  $r$  does not change too much during the passage of the lens. To increase the field close to the axis and to concentrate it into a small area, the coil is usually surrounded by an iron yoke. The focal length  $f_{sol}$  is given by

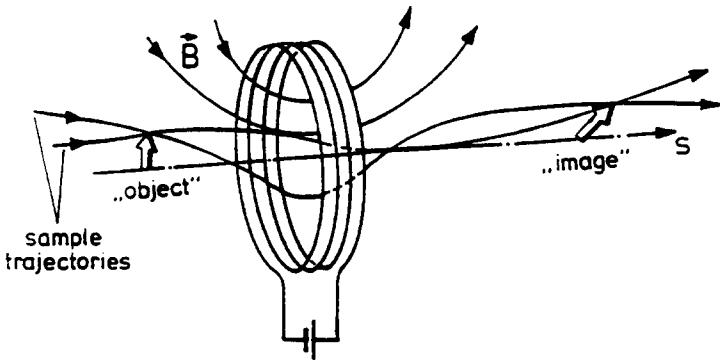


Figure 8: Particle trajectories and field lines in a “thin” lens formed by the solenoidal field of a coil (according to Bergmann/Schäfer: Optik)

$$\frac{1}{f_{sol}} = \int \left( \frac{eB_s}{2p} \right)^2 ds \quad (2.3)$$

In contrast to optical lenses, the image is rotated with respect to the object. As seen from Eq. (2.3),  $f_{sol}$  increases with the *square* of the momentum  $p$ . Therefore a solenoid lens is effective for small momenta only. At  $p \gg 1 \text{ MeV}/c$ , a quadrupole magnet is a much more effective lens, see the next section.

### c) Quadrupole magnet

Quadrupole magnets have four iron pole shoes with hyperbolic contour (Fig. 9).

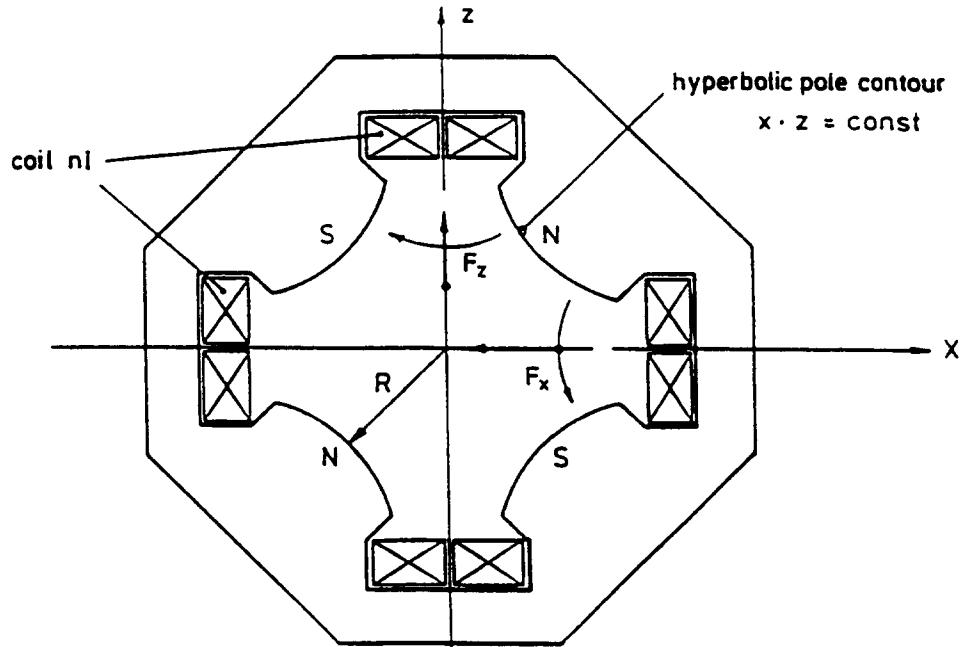


Figure 9: Cross-section of a quadrupole magnet. (Figs. 9,10,12 from K. Wille, Maria Laach lectures.)

With the polarity shown, the horizontal component of the Lorentz force on a positively charged

particle, moving into the plane of the drawing, is directed towards the axis, the vertical component is directed away from the axis. The magnet shown is thus horizontally focusing, vertically defocusing. The opposite holds when the current direction, the particle charge or its direction of motion is reversed.

The field is linear in the deviation from the axis:

$$B_z = -gx, \quad B_x = -gz \quad (2.4)$$

In the air space of the magnet which contains neither iron nor current conductors we have the Maxwell equation

$$\nabla \times \mathbf{B} = 0$$

Here the field can be written as the gradient of a potential

$$\mathbf{B} = -\nabla V \text{ with } V(x, z) = gxz \quad (2.5)$$

The equipotential lines are the hyperbolae  $xz = \text{const.}$  The field lines are perpendicular to them. If the relative permeability of the iron is large,  $\mu_r \gg 1$ , iron pole shoes with hyperbolic contour generate a rather pure quadrupole field (2.4).

The gradient  $g$  and the current  $I$  in the coils can be related by the integral theorem

$$\oint \mathbf{H} \cdot d\mathbf{s} = nI.$$

The path of integration is shown in Fig. 10.

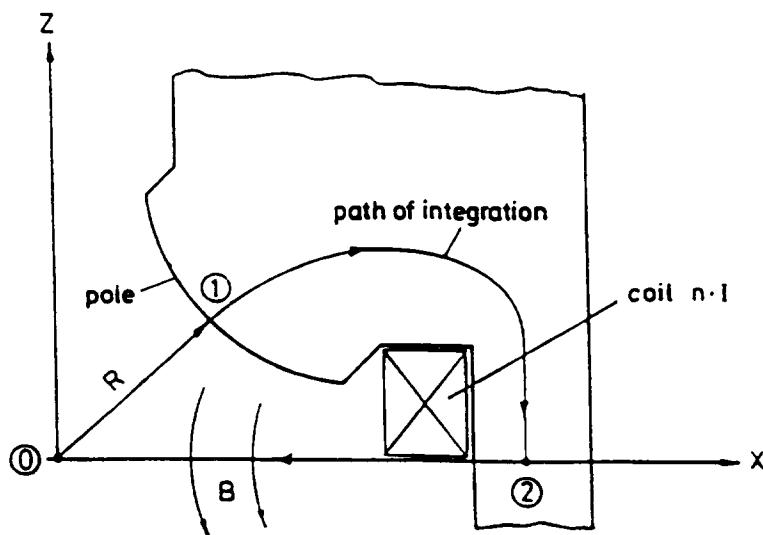


Figure 10: Path of integration used to compute the quadrupole gradient as a function of the current

$$nI = \oint \mathbf{H} \cdot d\mathbf{s} = \int_0^R H(r) dr + \int_1^2 \mathbf{H}_E \cdot d\mathbf{s} + \int_2^0 \mathbf{H} \cdot d\mathbf{s}$$

On the first path  $H(r) = gr/\mu_0$ . The second integral is very small for  $\mu_r \gg 1$ . The third integral vanishes identically since  $\mathbf{H} \perp d\mathbf{s}$ . So we get in good approximation

$$nI = \frac{1}{\mu_0} \int_0^R g r dr \quad r = \sqrt{x^2 + z^2}$$

$$g = \frac{2\mu_0 n I}{R^2} \quad (2.6)$$

In analogy to the bending strength  $1/\rho$  of a dipole magnet (see Eq. (2.2)), it is convenient to relate the field gradient to its optical effect. To this end, the field gradient is normalized to the momentum of the particle, thus defining the quadrupole strength

$$k = \frac{eg}{p}$$

Numerically

$$k [\text{m}^{-2}] = 0.2998 \frac{g [\text{T}/\text{m}]}{p [\text{GeV}/c]} \quad (2.7)$$

If  $l$  denotes the length of the quadrupole, its focal length  $f$  is given by

$$\frac{1}{f} = k \cdot l \quad (2.8)$$

Generally, a lens with  $f \gg l$  is called a “thin lens” —irrespective of the absolute value of  $l$ . An interesting property of the quadrupole is that the horizontal force component depends only on the horizontal position and not on the vertical position of the particle trajectory. Similarly, the vertical component of the Lorentz force depends only on the vertical position.

$$F_x = evB_z(x, z) = -evgx$$

$$F_z = -evB_x(x, z) = evgz \quad (2.9)$$

The important consequence is that in a so-called *linear machine*, containing only dipole and quadrupole fields, the horizontal and vertical betatron oscillations are completely *decoupled*. This is of great importance for an  $e^+e^-$  storage ring where the vertical betatron oscillation usually has a much smaller amplitude than the horizontal oscillation. This decoupling would be lost if quadrupole magnets were rotated by some angle around their longitudinal axis. With those rotated quadrupoles a flat beam would not remain flat.

Therefore: *Don't rotate quadrupoles unless you know what you are doing!*

#### d) “Synchrotron” magnet

As mentioned earlier, this magnet, sketched in Fig. 11, combines a dipole and a quadrupole magnet.

It can be considered as part of a quadrupole which is traversed at a distance  $d = B_0/g$  from the axis. The field can be derived from the potential

$$V(x, z) = -B_0 z + g \cdot xz \quad (2.10)$$

The bending strength  $1/\rho = eB_0/p$  and the quadrupole strength  $k = eg/p$  can be combined into the dimensionless “field index”  $n$ , see Eq. (1.5)

$$n = k \cdot \rho^2 \quad (2.11)$$

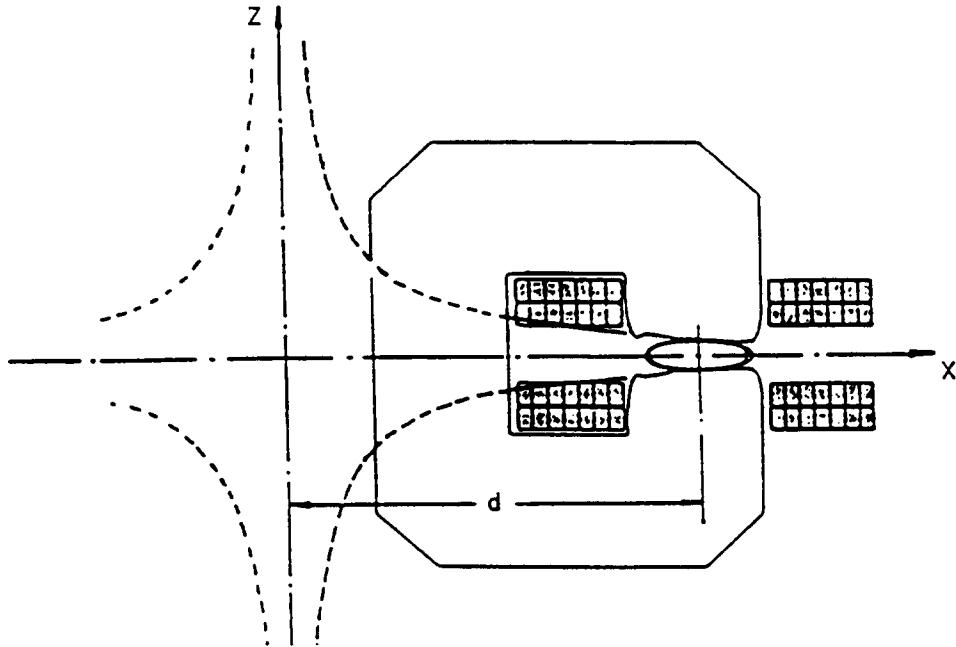


Figure 11: Cross-section of a horizontally focusing synchrotron magnet (from K. Steffen, Orsay lectures [4]).

e) Sextupole magnets

The focal length of a quadrupole depends on the particle momentum. Sextupole magnets (Fig. 12) are used to correct the resulting "chromatic" errors. A sextupole generates a nonlinear field

$$\begin{aligned} B_z &= \frac{1}{2}g'(x^2 - z^2) \\ B_x &= g'x \cdot z \end{aligned} \tag{2.12}$$

The field can be written as gradient of the potential

$$V(x, z) = -\frac{1}{2}g' \left( x^2 z - \frac{1}{3}z^3 \right)$$

A momentum-independent sextupole strength is defined by

$$m = \frac{eg'}{p_0}, \quad m [\text{m}^{-3}] = 0.2998 \frac{g' [\text{T/m}^2]}{p_0 [\text{GeV}/c]} \tag{2.13}$$

$g'$  is related to the current in the coils by

$$g' = 6\mu_0 n I / R^3$$

f) General multipole expansion

In the literature you may find various kinds of multipole expansions for the guide field: expansion of the magnetic field  $\mathbf{B}$ , expansion of the magnetic vector potential  $\mathbf{A}$ , expansion of the scalar potential, and using cylindrical as well as cartesian coordinates. People use whatever is most

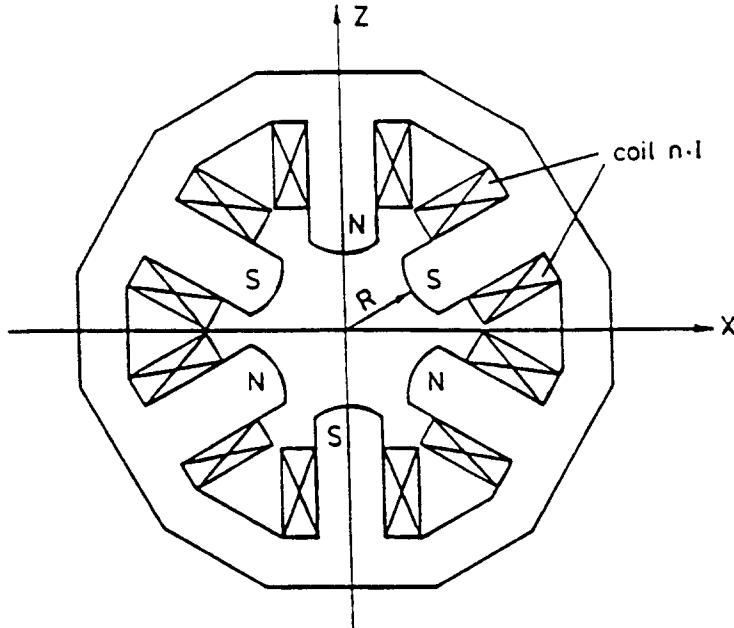


Figure 12: Sextupole magnet

appropriate for the specific problem, but the physics content is always the same as will be shown in the following.

In this paragraph we denote the vertical coordinate by  $y$  instead of  $z$  because we want to keep the conventional notation  $z = x + iy$  for complex numbers.

The length of modern accelerator magnets is usually much larger than their bore radius. The end field contribution is then rather small and the magnetic field has to a good approximation only transverse components. (This is of course not the case for the large solenoids in the experimental areas which need a special treatment. The same applies for wigglers and undulators which are special magnets for generating synchrotron radiation.)

For two-dimensional fields one can apply the theory of analytic functions. From

$$\operatorname{div} \mathbf{B} = 0$$

it follows that a vector potential  $\mathbf{A}$  exists such that

$$\mathbf{B} = \operatorname{rot} \mathbf{A} \quad (2.14)$$

Because of the transversality of the field, the vector potential has only a component  $A_s$  in the longitudinal direction  $s$ . In vacuum, for example inside the beam pipe, we have furthermore

$$\operatorname{rot} \mathbf{B} = 0$$

This implies that  $\mathbf{B}$  can also be written as the gradient of a scalar potential  $V$ :

$$\mathbf{B} = -\operatorname{grad} V \quad (2.15)$$

Combining both equations (2.14, 2.15) we get:

$$B_x = -\frac{\partial V}{\partial x} = \frac{\partial A_s}{\partial y} \quad B_y = -\frac{\partial V}{\partial y} = -\frac{\partial A_s}{\partial x} \quad (2.16)$$

Now we define a complex potential function of  $z = x + iy$  by

$$\tilde{A}(z) = A_s(x, y) + iV(x, y) \quad (2.17)$$

The equations (2.16) are just the Cauchy-Riemann conditions for the real and imaginary part of an analytic function. So the complex potential is an analytic function and can be expanded in a power series

$$\tilde{A}(z) = \sum_{n=0}^{\infty} \kappa_n z^n \quad \kappa_n = \lambda_n + i\mu_n \quad (2.18)$$

with  $\lambda_n, \mu_n$  being real constants.

From complex analysis we know that this series expansion converges for all  $z$  inside a circle  $|z| < r_c$ . The radius of convergence  $r_c$  is the closest distance between the origin of the expansion and the iron yoke or the coil where the Eqs. (2.16) break down and  $\tilde{A}(z)$  is no more analytic, see Fig. 13.

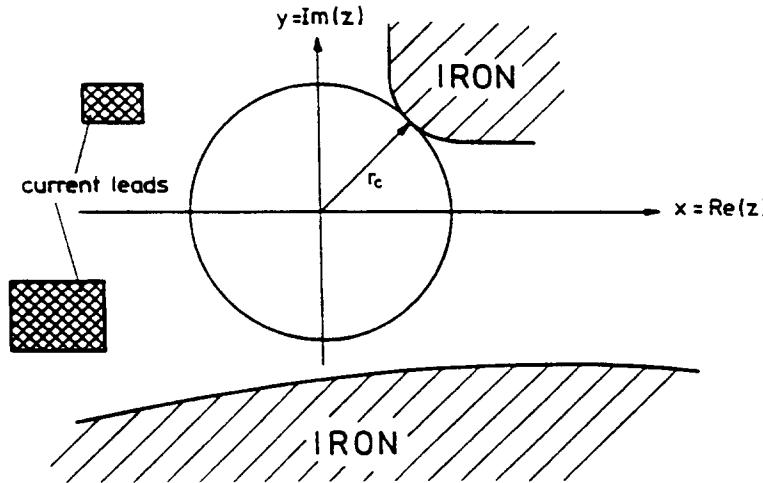


Figure 13: The multipole expansion with respect to  $z = 0$  is only valid inside the circle  $r_c$  (radius of convergence).

#### Cylindrical coordinate representation

For superconducting magnets, it is practical to express the field in cylindrical coordinates  $(r, \varphi, s)$ , see Fig. 14:

$$x = r \cos \varphi \quad y = r \sin \varphi \quad z^n = r^n \cdot e^{in\varphi} = r^n (\cos n\varphi + i \sin n\varphi) \quad (2.19)$$

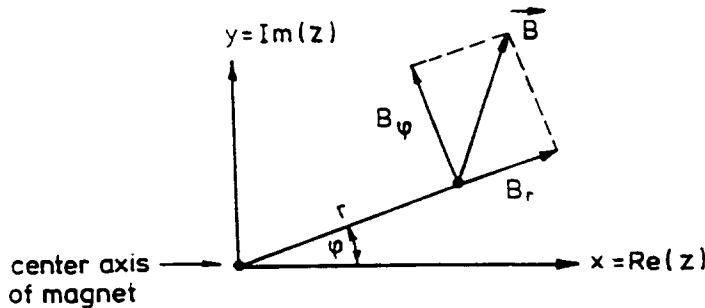


Figure 14: Cylindrical coordinate system used in the multipole expansion

The scalar potential is given by the imaginary part of Eq. (2.18)

$$V(r, \varphi) = \sum_{n=0}^{\infty} (\mu_n \cos n\varphi + \lambda_n \sin n\varphi) r^n \quad (2.20)$$

Similarly, we get from the real part of Eq. (2.18)

$$A_s(r, \varphi) = \sum_{n=0}^{\infty} (\lambda_n \cos n\varphi - \mu_n \sin n\varphi) r^n \quad (2.21)$$

Taking the gradient of  $-V(r, \varphi)$ , we get the multipole expansion of the azimuthal and radial field components, respectively

$$\begin{aligned} B_\varphi &= -\frac{1}{r} \frac{\partial V}{\partial \varphi} = -\sum_{n=1}^{\infty} n(\lambda_n \cos n\varphi - \mu_n \sin n\varphi) r^{n-1} \\ B_r &= -\frac{\partial V}{\partial r} = -\sum_{n=1}^{\infty} n(\mu_n \cos n\varphi + \lambda_n \sin n\varphi) r^{n-1} \end{aligned}$$

Now it is convenient to define a ‘reference radius’  $r_0$  for the multipole expansion and to denote the magnitude of the main field component of the magnet in question by  $B_{main}$ . A useful choice for  $r_0$  is the largest conceivable deviation of beam particles from the design orbit (25 mm in HERA, that is the inner radius of the beam pipe). Furthermore we introduce the ‘normal’ multipole coefficients  $b_n$  and the ‘skew’ coefficients  $a_n$  by

$$b_n = -\frac{n\lambda_n}{B_{main}} r_0^{n-1} \quad a_n = +\frac{n\mu_n}{B_{main}} r_0^{n-1} \quad (2.22)$$

Then the multipole expansions read (note that  $a_0, b_0$  are set to zero as they don’t contribute to the magnetic field)

$$V(r, \varphi) = -B_{main} r_0 \sum_{n=1}^{\infty} \left( -\frac{a_n}{n} \cos n\varphi + \frac{b_n}{n} \sin n\varphi \right) \left( \frac{r}{r_0} \right)^n \quad (2.23)$$

$$A_s(r, \varphi) = -B_{main} r_0 \sum_{n=1}^{\infty} \left( \frac{b_n}{n} \cos n\varphi + \frac{a_n}{n} \sin n\varphi \right) \left( \frac{r}{r_0} \right)^n \quad (2.24)$$

$$B_\varphi(r, \varphi) = B_{main} \sum_{n=1}^{\infty} (b_n \cos n\varphi + a_n \sin n\varphi) \left( \frac{r}{r_0} \right)^{n-1} \quad (2.25)$$

$$B_r(r, \varphi) = B_{main} \sum_{n=1}^{\infty} (-a_n \cos n\varphi + b_n \sin n\varphi) \left( \frac{r}{r_0} \right)^{n-1} \quad (2.26)$$

Remember that these multipole expansions are only valid within a circle of radius  $r_c$  containing neither iron nor current! For an ideal  $2n$ -pole magnet we have  $b_n = 1$  and all other  $a_n, b_n = 0$ . We call

$n = 1$	Dipole
$n = 2$	Quadrupole
$n = 3$	Sextupole
$n = 4$	Octupole
$n = 5$	Decapole
$n = 6$	Dodecapole or 12 – pole

It is instructive to consider  $B_\varphi + iB_r$ :

$$B_\varphi + iB_r = B_{main} \sum_{n=1}^{\infty} \left( \frac{r}{r_0} \right)^{n-1} [b_n(\cos n\varphi + i \sin n\varphi) - ia_n(\cos n\varphi + i \sin n\varphi)]$$

$$B_\varphi + iB_r = B_{\text{main}} \sum_{n=1}^{\infty} \left(\frac{r}{r_0}\right)^{n-1} (b_n - ia_n) e^{in\varphi} \quad (2.27)$$

Thus

$$(|\mathbf{B}|)_n = \left( \sqrt{B_r^2 + B_\varphi^2} \right)_n = B_{\text{main}} \left( \frac{r}{r_0} \right)^{n-1} \sqrt{a_n^2 + b_n^2} \quad (2.28)$$

i.e. the magnitude of the  $2n$  pole field component *does not* depend on the azimuth and scales with the  $(n-1)$ th power of  $r$ . Equation (2.28) also illustrates a simple interpretation of the fractional multipole field coefficients  $a_n$ ,  $b_n$ : They are just the relative field contribution of the  $n$ th multipole to the main field at the reference radius  $r_0$ . This is the reason why the coefficients  $\lambda_n$ ,  $\mu_n$  have been normalized with Eqs. (2.22).

Conventional accelerator magnets with iron pole shoes are limited to dipole fields of about 2 T and quadrupole gradients of about 20 T/m. Significantly higher values ( $> 6$  T,  $> 100$  T/m) are possible with superconducting magnets. In these magnets the field distribution is entirely determined by the conductor arrangement and the coils have to be built with extreme accuracy to keep field distortions below the required level of  $10^{-4}$ . Figure 15a shows schematically the layout of a superconducting dipole.

Ideally, the current as a function of the azimuthal angle  $\varphi$  should follow a  $\cos \varphi$ -distribution to generate a pure dipole field and a  $\cos 2\varphi$  ( $\cos 3\varphi$ ) distribution for a quadrupole (sextupole) field (Figs. 15b,c,d). Since these ideal distributions are technically difficult to realize one approximates them by an arrangement of current shells. The cylindrical coordinate representation is particularly useful for magnet design from current shells. Figure 16 shows a cross section through the HERA dipole coil.

Another application of the cylindrical coordinate representation is the technique of measurement of the multipole components with a coil rotating in the field: The  $n$ th Fourier component of the induced voltage is proportional to  $\sqrt{a_n^2 + b_n^2}$  while its phase is related to  $a_n/b_n$ .

In a good dipole or quadrupole magnet the unwanted multipole coefficients  $a_n$ ,  $b_n$  are typically a few  $10^{-4}$  or less.

Finally it is noted that within the cylindrical coordinate representation one easily understands which multipole components are forbidden if specific symmetry properties of the field are assumed. For instance, for a quadrupole with perfect constructional symmetry only *odd* harmonics of the 4-pole are allowed. Or, as another example, if mirror symmetry with respect to the  $x-s$  plane is assumed, all skew components are forbidden since  $B_\varphi$  must behave purely *cos-like*. A similar reasoning shows that any normal  $2n$ -pole magnet transforms into a skew  $2n$ -pole magnet if rotated by  $\pi/2n$ .

### Cartesian coordinates

In cartesian coordinates, Eq. (2.18) reads

$$A_s(x, y) + iV(x, y) = \sum_{n=0}^{\infty} \kappa_n z^n = \sum_{n=0}^{\infty} (\lambda_n + i\mu_n)(x + iy)^n \quad (2.29)$$

Separation of real and imaginary part and use of Eq. (2.22) yields

$$\begin{aligned} A_s(x, y) &= \Re \sum_{n=0}^{\infty} \kappa_n z^n = -B_{\text{main}} [b_1 x + a_1 y + \frac{b_2}{2r_0}(x^2 - y^2) + \frac{a_2}{r_0}xy + \frac{b_3}{3r_0^2}(x^3 - 3xy^2) + \\ &\quad + \frac{a_3}{3r_0^2}(3x^2y - y^3) + \frac{b_4}{4r_0^3}(x^4 - 6x^2y^2 + y^4) + \frac{a_4}{r_0^3}(x^3y - xy^3) \pm \dots] \end{aligned} \quad (2.30)$$

$$V(x, y) = \Im \sum_{n=0}^{\infty} \kappa_n z^n = B_{\text{main}} [a_1 x - b_1 y + \frac{a_2}{2r_0}(x^2 - y^2) - \frac{b_2}{r_0}xy + \frac{a_3}{3r_0^2}(x^3 - 3xy^2) -$$

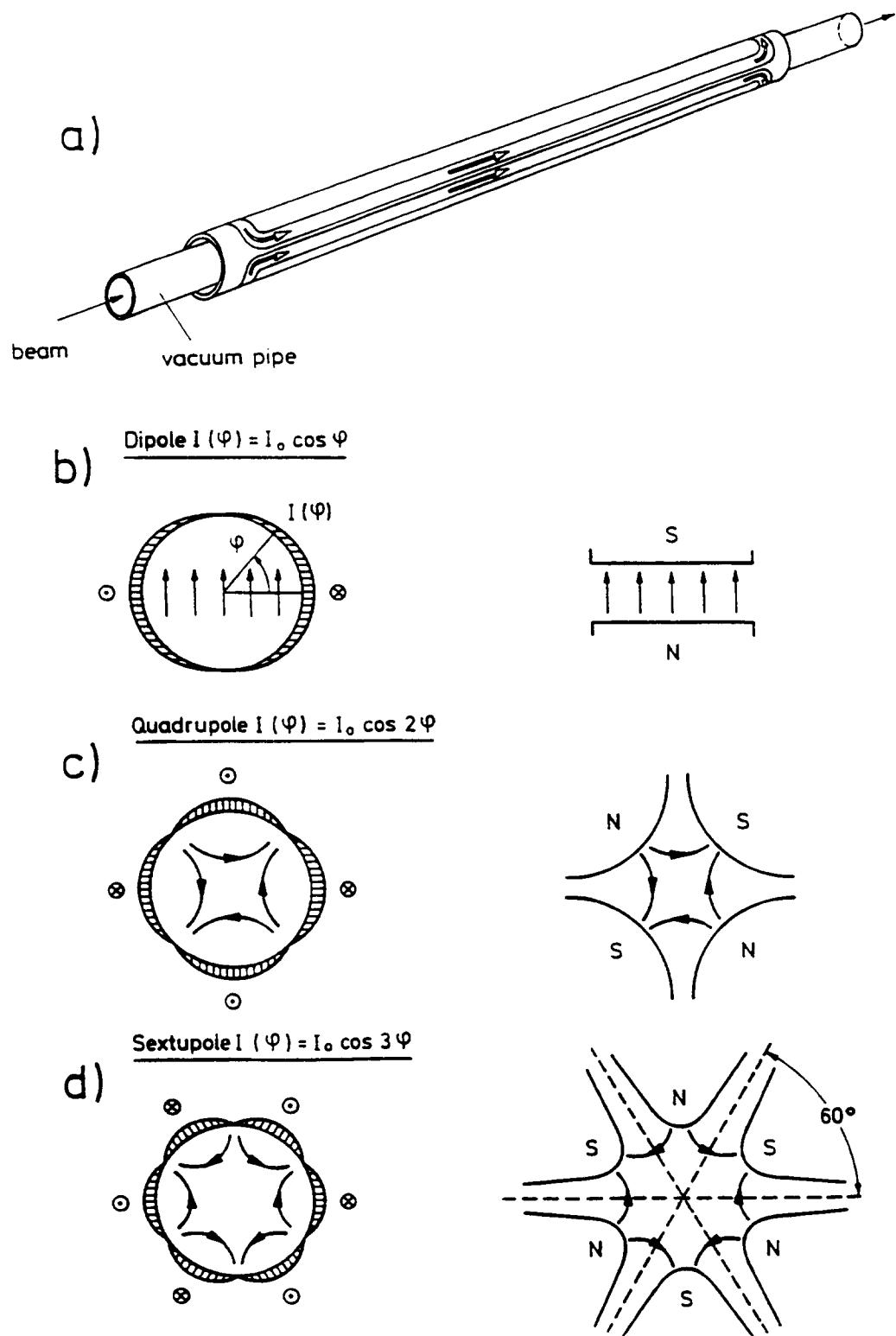


Figure 15: a) Schematic view of a superconducting dipole  
 b, c, d) current distributions for pure dipole, quadrupole and sextupole fields  
 and the corresponding yoke profiles of conventional magnets.

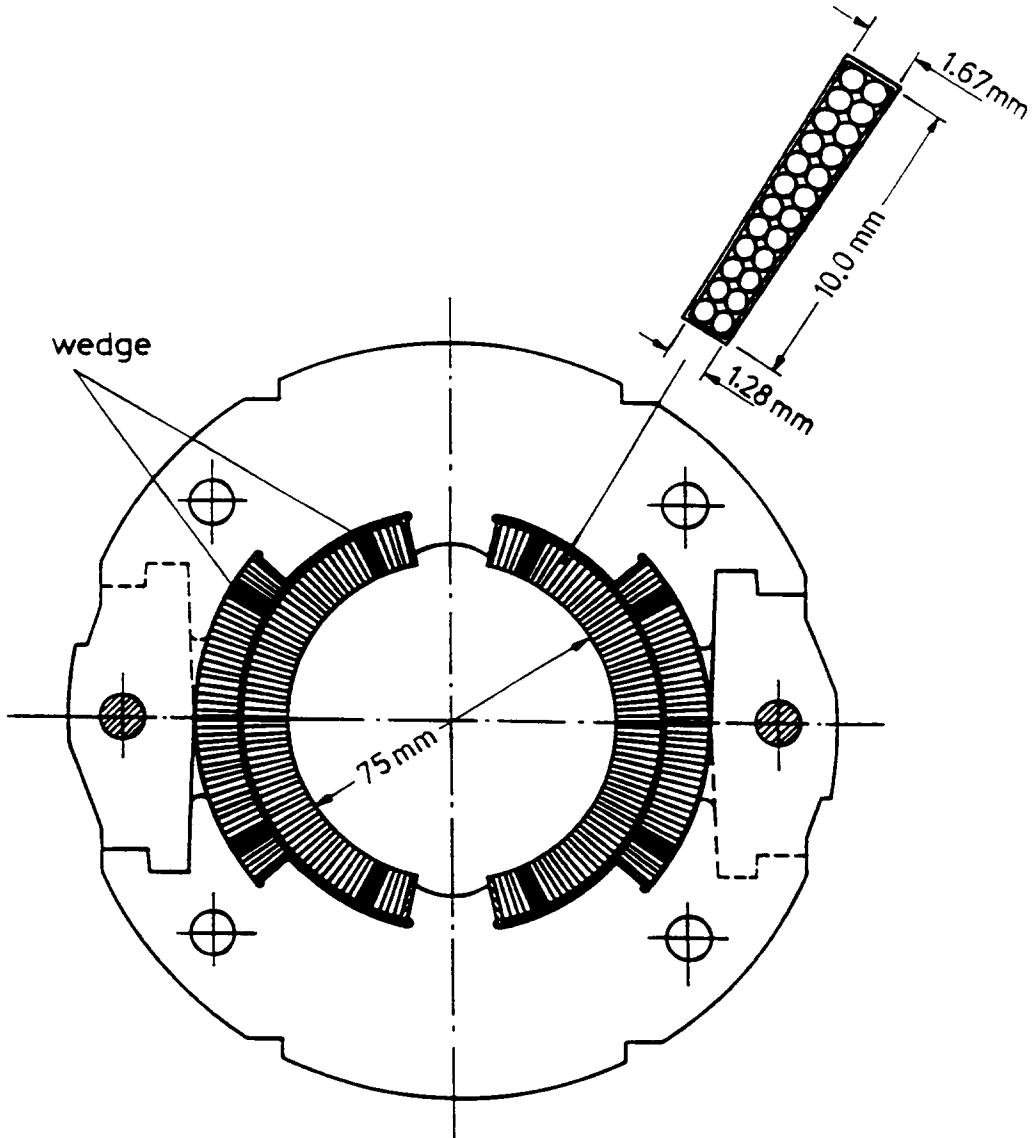


Figure 16: Cross section of the superconducting HERA dipole coil. The inner current shell has 64 windings, the outer shell 40. The limiting angles of the shells and the longitudinal wedges are chosen such that the computed higher multipole fields are less than  $10^{-4}$  of the dipole field. The nominal field of 4.68 T is achieved with a current of 5025 A. The coil is confined by precise aluminum clamps which sustain the magnetic forces and define the exact geometry. The clamps are surrounded with a cylindrical iron yoke (not shown) contributing 22 % to the field. The cross section of the superconducting cable is shown in an enlarged view.

$$-\frac{b_3}{3r_0^2}(3x^2y - y^3) + \frac{a_4}{4r_0^3}(x^4 - 6x^2y^2 + y^4) - \frac{b_4}{r_0^3}(x^3y - xy^3) \pm \dots] \quad (2.31)$$

To get the cartesian components of the magnetic field we now have to take the gradient of  $-V(x, y)$  in cartesian coordinates, see Eq. (2.16)

$$\begin{aligned} B_x(x, y) = -\frac{\partial V}{\partial x} &= B_{main}[-a_1 + \frac{b_2}{r_0}y - \frac{a_2}{r_0}x - \frac{a_3}{r_0^2}(x^2 - y^2) + \frac{b_3}{r_0^2}2xy - \\ &\quad - \frac{a_4}{r_0^3}(x^3 - 3xy^2) + \frac{b_4}{r_0^3}(3x^2y - y^3) \pm \dots] \end{aligned} \quad (2.32)$$

$$\begin{aligned} B_y(x, y) = -\frac{\partial V}{\partial y} &= B_{main}[b_1 + \frac{a_2}{r_0}y + \frac{b_2}{r_0}x + \frac{a_3}{r_0^2}2xy + \frac{b_3}{r_0^2}(x^2 - y^2) + \\ &\quad + \frac{a_4}{r_0^3}(3x^2y - y^3) + \frac{b_4}{r_0^3}(x^3 - 3xy^2) \pm \dots] \end{aligned} \quad (2.33)$$

Another useful combination of Eqs. (2.32) and (2.33) is

$$\begin{aligned} B_y + iB_x &= -\frac{\partial}{\partial x}(A_s + iV) = -\sum_{n=1}^{\infty} n(\lambda_n + i\mu_n)(x + iy)^{n-1} \\ B_y + iB_x &= B_{main} \sum_{n=1}^{\infty} (b_n - ia_n) \left( \frac{x}{r_0} + i \frac{y}{r_0} \right)^{n-1} \end{aligned}$$

Here are two applications of the cartesian representation of multipoles:

- If the motion of particles is described in cartesian coordinates, the contribution of each individual multipole to the equation of motion is easily identified. As stated before, the coefficients  $b_n$  are called the “normal” multipole coefficients,  $a_n$  are the “skew” coefficients. In magnets containing normal coefficients only, a flat beam (i.e. no vertical extension) remains flat forever, since for  $y \equiv 0$  there is  $B_x \equiv 0$ , i.e. no vertical force. Thus, there is no coupling of horizontal motion into the vertical.
- Equation (2.31) is useful in conventional lens design work with iron pole shoes. It describes the pole contours of dipole ( $n = 1$ ), quadrupole ( $n = 2$ ), sextupole ( $n = 3$ ), octupole ( $n = 4$ ), etc., magnets, because the pole contour is a line of constant magnetic potential. The pole contour of a normal quadrupole ( $b_2$ ), for instance, is given by the hyperbola  $x \cdot y = const$  (see Fig. 9 and Eq. (2.5)).

Finally, we show explicitly the field distribution of the most important multipole components:

Normal dipole ( $n = 1$ ):	$b_1 \cdot B_{\text{main}} = B_{\text{vert}}$	(horizontally bending)
$B_\varphi(r, \varphi) = B_{\text{vert}} \cdot \cos \varphi$	$B_r(r, \varphi) = B_{\text{vert}} \cdot \sin \varphi$	
$B_x(x, y) = 0$	$B_y(x, y) = B_{\text{vert}}$	
Skew dipole ( $n = 1$ ):	$a_1 \cdot B_{\text{main}} = B_{\text{hor}}$	(vertically bending)
$B_\varphi(r, \varphi) = B_{\text{hor}} \cdot \sin \varphi$	$B_r(r, \varphi) = -B_{\text{hor}} \cdot \cos \varphi$	
$B_x(x, y) = -B_{\text{hor}}$	$B_y(x, y) = 0$	
Normal quadrupole ( $n = 2$ ):	$b_2 \cdot B_{\text{main}} = -g \cdot r_0$	(where $g$ is the gradient)
$B_\varphi(r, \varphi) = -gr \cos 2\varphi$	$B_r(r, \varphi) = -gr \sin 2\varphi$	
$B_x(x, y) = -gy$	$B_y(x, y) = -gx$	
Skew quadrupole ( $n = 2$ ):	$a_2 \cdot B_{\text{main}} = -g \cdot r_0$	(quadrupole rotated by 45°)
$B_\varphi(r, \varphi) = -gr \sin 2\varphi$	$B_r(r, \varphi) = gr \cos 2\varphi$	
$B_x(x, y) = gx$	$B_y(x, y) = -gy$	
Normal sextupole ( $n = 3$ ):	$b_3 \cdot B_{\text{main}} = \frac{1}{2}g' \cdot r_0^2$	
$B_\varphi(r, \varphi) = \frac{1}{2}g'r^2 \cos 3\varphi$	$B_r(r, \varphi) = \frac{1}{2}g'r^2 \sin 3\varphi$	
$B_x(x, y) = g'xy$	$B_y(x, y) = \frac{1}{2}g'(x^2 - y^2)$	

### 3 PARTICLE MOTION IN A CIRCULAR ACCELERATOR

#### 3.1 Trajectory equations

We consider first a machine with combined-function magnets and neglect the drift spaces between the magnets. The design orbit is then a circle. It is easy to generalize the results to machines with separated-function magnets.

The design orbit is assumed to be in the horizontal plane. We use cylindrical coordinates,  $r, \Theta, z$ . The dipole field is assumed to point in the positive  $z$  direction. Then positively charged particles move in a clockwise direction, negatively charged particles counterclockwise (see Fig. 17).

Let us follow the motion of an electron. Together with the particle a system of three unit vectors  $\mathbf{u}_r$ ,  $\mathbf{u}_\Theta$  and  $\mathbf{u}_z$  is moved around. The radius vector of the particle is

$$\mathbf{R} = \mathbf{R}_0 + r\mathbf{u}_r + z\mathbf{u}_z \quad \mathbf{R}_0 = \text{const.}$$

Now for a small  $d\Theta$

$$d\mathbf{u}_r = d\Theta \mathbf{u}_\Theta, \quad d\mathbf{u}_\Theta = -d\Theta \mathbf{u}_r, \quad d\mathbf{u}_z = 0$$

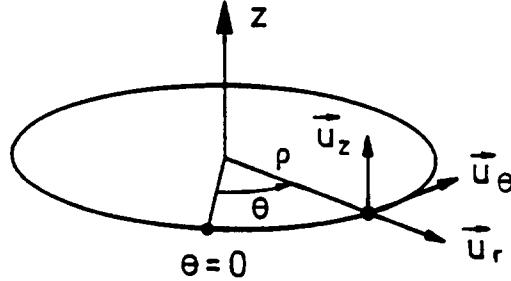


Figure 17: Coordinate system for particle motion in a circular accelerator

So

$$\begin{aligned}\dot{\mathbf{R}} &= \dot{r}\mathbf{u}_r + r\dot{\theta}\mathbf{u}_\theta + \dot{z}\mathbf{u}_z \\ &= \dot{r}\mathbf{u}_r + r\dot{\theta}\mathbf{u}_\theta + \dot{z}\mathbf{u}_z\end{aligned}$$

The acceleration is

$$\ddot{\mathbf{R}} = (\ddot{r} - r\dot{\theta}^2)\mathbf{u}_r + (2\dot{r}\dot{\theta} + r\ddot{\theta})\mathbf{u}_\theta + \ddot{z}\mathbf{u}_z$$

Now

$$m\ddot{\mathbf{R}} = -e\mathbf{v} \times \mathbf{B} = -e \left[ (r\dot{\theta}B_z - \dot{z}B_\theta) \mathbf{u}_r + (\dot{z}B_r - \dot{r}B_z) \mathbf{u}_\theta + (\dot{r}B_\theta - r\dot{\theta}B_r) \mathbf{u}_z \right] \quad (3.1)$$

In the following we assume  $B_\theta = 0$  and obtain

$$\begin{aligned}m(\ddot{r} - r\dot{\theta}^2) &= -er\dot{\theta}B_z(r, z, \theta) \\ m\ddot{z} &= er\dot{\theta}B_r(r, z, \theta)\end{aligned} \quad (3.2)$$

In the special case of combined-function magnets the field components derived from the potential (2.10) are

$$B_z = B_0 - gx, \quad B_r = B_x = -gz \quad (3.3)$$

Inserting (3.3) into (3.2) and using  $r = \rho + x$  with  $\rho = \text{const}$ , we get

$$\begin{aligned}m(\ddot{x} - r\dot{\theta}^2) &= -er\dot{\theta}(B_0 - gx) \\ m\ddot{z} &= -er\dot{\theta}gz\end{aligned}$$

$v_\theta = r\dot{\theta}$  is the azimuthal component of the particle velocity. It is much larger than the transverse components  $v_r$  and  $v_z$ , so  $v_\theta = r\dot{\theta} \approx v$ . At this point it is convenient to replace the time variable  $t$  by the arc length  $s$  along the design orbit

$$s = vt, \quad \ddot{x} = v^2 x'', \quad x'' = \frac{d^2x}{ds^2}$$

$$x'' = \frac{1}{r} - \frac{eB_0}{mv} + \frac{eg}{mv}x$$

$$z'' = -\frac{eg}{mv}z$$

Now  $mv = p = p_0 \left(1 + \frac{\Delta p}{p_0}\right)$  where  $p_0$  is the design momentum.

Further  $\frac{1}{r} \approx \frac{1}{\rho} \left(1 - \frac{x}{\rho}\right)$  for  $x \ll \rho$  and  $p_0 = eB_0$ ,  $\frac{1}{\rho} = \frac{eB_0}{p_0}$ ,  $k = \frac{eg}{p_0}$

To first order in the small quantities  $x$  and  $\frac{\Delta p}{p_0}$  we obtain

$$x'' - \left(k - \frac{1}{\rho^2}\right)x = \frac{1}{\rho} \frac{\Delta p}{p_0}$$

$$z'' + kz = 0$$

(3.4)

These are the basic equations for the particle trajectory  $x(s), z(s)$  in linear approximation.

It is easy to see that the equations (3.4) are also valid for a separated-function machine, provided we define  $\rho$  as the local radius of curvature and  $g$  as the local gradient. In a dipole magnet the design orbit is part of a circle. The gradient is zero, so the equations hold with  $k = 0$ . In a quadrupole the central orbit is a straight line. The force on the electron is

$$m\ddot{x} = -evB_z = evgx$$

$$m\ddot{z} = evB_x = -evgz$$

$$\text{so } x'' - kx = 0$$

$$z'' + kz = 0 \quad k = \frac{eg}{p_0} \approx \frac{eg}{mv}$$

So Eqs. (3.4) are valid with  $\frac{1}{\rho} = 0$ .

The term  $\frac{1}{\rho}x$  in (3.4) describes the “weak focusing” of a bending magnet. In the very large high-energy accelerators like the CERN SPS or HERA it can be neglected in comparison with the “strong focusing” term  $k$  given by the quadrupoles.

Example HERA proton ring:  $k = 0.032\text{m}^{-2}$ ,  $\frac{1}{\rho} = 2.9 \cdot 10^{-6}\text{m}^{-2}$ .

We want to point out that even in the case of purely linear magnetic fields the equations of motion are intrinsically nonlinear due to the curved coordinate system. Especially for accelerators with small bending radius, nonlinear corrections to Eq. (3.4) have to be taken into account.

### 3.2 Solution of trajectory equations in terms of principal trajectories

In general, the bending strength  $\frac{1}{\rho}$  and the focusing strength  $k$  are functions of the path length  $s$  along the reference orbit. The equation for the horizontal motion can be written as

$$x'' + K(s)x = \frac{1}{\rho} \cdot \frac{\Delta p}{p_0} \quad (K(s) = -k(s) + \frac{1}{\rho^2(s)}) \quad (3.5)$$

The general solution  $x(s)$  is the sum of the complete solution  $x_h$  of the homogeneous equation and a particular solution  $x_i$  of the inhomogeneous equation

$$x(s) = x_h(s) + x_i(s)$$

with

$$\begin{aligned} x_h'' + K(s)x_h &= 0 \\ x_i'' + K(s)x_i &= \frac{1}{\rho} \frac{\Delta p}{p_0} \end{aligned}$$

Since  $\Delta p/p_0$  is assumed to be constant it is obvious that if  $x_i$  is a solution for a given  $\Delta p/p_0$ ,  $n \cdot x_i$  will be a solution for  $n \cdot \Delta p/p_0$ . Therefore it is useful to normalize  $x_i$  with respect to  $\Delta p/p_0$ :

$$D(s) = \frac{x_i}{\Delta p/p_0}$$

The general solution now reads

$$x(s) = C(s)x_0 + S(s)x'_0 + D(s)\frac{\Delta p}{p_0} \quad (3.6)$$

Here  $x_0$ ,  $x'_0$  are the initial values of  $x_h(s)$  and  $x'_h(s)$  at  $s = s_0$  and  $C(s)$  and  $S(s)$  are two independent solutions of the homogeneous equation

$$C'' + K(s)C = 0, \quad S'' + K(s)S = 0$$

For  $C(s)$  and  $S(s)$  to be linearly independent, the Wronski determinant  $W$  has to meet the condition

$$W = \begin{vmatrix} C & S \\ C' & S' \end{vmatrix} \neq 0$$

The derivative of the Wronskian vanishes identically

$$\frac{d}{ds}(CS' - SC') = \frac{d}{ds}W = CS'' - SC'' = -K(CS - SC) = 0$$

So the value of  $W$  is determined everywhere by the initial conditions at  $s = s_0$ . We choose

$$C_0 = 1, C'_0 = 0; \quad S_0 = 0, S'_0 = 1 \quad (3.7)$$

The solutions satisfying these initial conditions are called the “Cosinelike” and the “Sinelike” trajectories and they result in  $W = 1$ .

$D(s)$  is the “Dispersion” trajectory, defined as a particular solution of the following inhomogeneous equation

$$D''(s) + K(s)D(s) = \frac{1}{\rho(s)} \quad (3.8)$$

It describes the momentum-dependent part of the motion. Here we require the initial conditions

$$D_0 = D'_0 = 0$$

because we assume (quite arbitrarily) that particles of different energy are not separated in space at the beginning.

**Remark:** The closed dispersion trajectory defined in chapter 5 satisfies different, namely periodic, boundary conditions.

$x(s)$  and  $x'(s)$  are related to their initial values by a linear transformation

$$\begin{pmatrix} x \\ x' \end{pmatrix}_s = \begin{pmatrix} C & S \\ C' & S' \end{pmatrix} \begin{pmatrix} x \\ x' \end{pmatrix}_{s_0} + \frac{\Delta p}{p_0} \begin{pmatrix} D \\ D' \end{pmatrix} \quad (3.9)$$

Equation (3.9) can also be written as

$$\begin{pmatrix} x \\ x' \\ \frac{\Delta p}{p_0} \end{pmatrix}_s = \begin{pmatrix} C & S & D \\ C' & S' & D' \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ x' \\ \frac{\Delta p}{p_0} \end{pmatrix}_{s_0} \quad (3.10)$$

The dispersion trajectory can be calculated from the cosinelike and the sinelike trajectories:

$$D(s) = S(s) \int_{s_0}^s \frac{1}{\rho(t)} C(t) dt - C(s) \int_{s_0}^s \frac{1}{\rho(t)} S(t) dt \quad (3.11)$$

To prove this relation we show that  $D(s)$  as given by (3.11) fulfills the equation (3.8) with the initial conditions  $D_0 = D'_0 = 0$ .

$$\begin{aligned} D' &= S' \int \frac{1}{\rho} C dt - C' \int \frac{1}{\rho} S dt \\ D'' &= S'' \int \frac{C}{\rho} dt - C'' \int \frac{S}{\rho} dt + \underbrace{\frac{1}{\rho} (CS' - SC')}_1 \\ D''(s) &= -K(s)D(s) + \frac{1}{\rho(s)} \end{aligned}$$

The vertical motion is of course described by an equation like (3.9) but with the last term missing.

### 3.3 Transformation matrices of accelerator magnets

The matrix notation of the solution of the equations of motion is particularly useful if  $K(s)$  is piecewise constant, because with  $K = \text{const}$  the matrix elements can be expressed analytically. The solution for the complete “lattice” of optical elements is then just the product of the individual matrices in the desired sequence, see Fig. 18. The transfer matrices of the most important magnets are discussed in the following. They may be used as building bricks to assemble the complete magnet lattice.

#### a) Combined-function magnet

Although not much used anymore, this magnet is treated first since dipole and quadrupole magnets can be considered as special cases of a synchrotron magnet.

We assume that the field is independent of  $s$  inside the magnet and drops abruptly to zero at the ends (hard-edge model). The principal trajectories  $C(s), S(s)$  solve the equation

$$y'' + Ky = 0 \quad \begin{cases} K = k \text{ for } y = z \\ K = -k + \frac{1}{\rho^2} \text{ for } y = x \end{cases}$$

With  $\varphi = s\sqrt{|K|}$  the solutions are inside the magnet

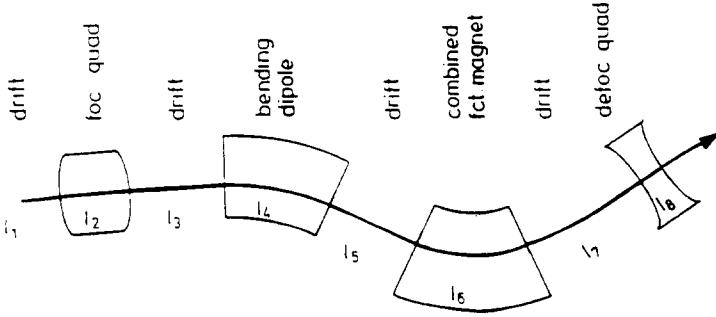


Figure 18: The complete transfer matrix of this sequence of magnetic elements is the matrix product  $M_{tot} = M_8 \cdot M_7 \cdot M_6 \cdot M_5 \cdot M_4 \cdot M_3 \cdot M_2 \cdot M_1$ . Each of the matrices  $M_1 \dots M_8$  describes a section with  $K(s) = \text{const}$ .

$$\begin{pmatrix} C & S \\ C' & S' \end{pmatrix} = \begin{pmatrix} \cos \varphi & \frac{1}{\sqrt{|K|}} \sin \varphi \\ -\sqrt{|K|} \sin \varphi & \cos \varphi \end{pmatrix} \quad \begin{array}{l} \text{for } K > 0 \\ (\text{focusing}) \end{array} \quad (3.12a)$$

$$\begin{pmatrix} C & S \\ C' & S' \end{pmatrix} = \begin{pmatrix} \cosh \varphi & \frac{1}{\sqrt{|K|}} \sinh \varphi \\ \sqrt{|K|} \sinh \varphi & \cosh \varphi \end{pmatrix} \quad \begin{array}{l} \text{for } K < 0 \\ (\text{defocusing}) \end{array} \quad (3.12b)$$

$$\begin{pmatrix} C & S \\ C' & S' \end{pmatrix} = \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix} \quad \begin{array}{l} \text{for } K = 0 \\ (\text{drift space}) \end{array} \quad (3.12c)$$

The determinant is indeed unity:  $CS' - SC' = 1$ .

The dispersion is for  $K > 0$

$$\begin{aligned} D(s) &= S \int_0^s \frac{C}{\rho} dt - C \int_0^s \frac{S}{\rho} dt \\ &= \frac{1}{\rho K} \sin^2 \varphi + \frac{1}{\rho K} \cos \varphi (\cos \varphi - 1) \\ D(s) &= \frac{1}{\rho K} (1 - \cos \varphi) \end{aligned}$$

For  $K < 0$ :  $D(s) = -\frac{1}{\rho |K|} (1 - \cosh \varphi)$

$$\begin{pmatrix} D \\ D' \end{pmatrix} = \begin{pmatrix} \frac{1}{\rho K} \cdot (1 - \cos \varphi) \\ \frac{1}{\rho \sqrt{|K|}} \cdot \sin \varphi \end{pmatrix} \quad \begin{array}{l} \text{for } K > 0 \\ \text{for } K < 0 \end{array} \quad (3.13a)$$

$$\begin{pmatrix} D \\ D' \end{pmatrix} = \begin{pmatrix} -\frac{1}{\rho |K|} \cdot (1 - \cosh \varphi) \\ \frac{1}{\rho \sqrt{|K|}} \cdot \sinh \varphi \end{pmatrix} \quad \begin{array}{l} \text{for } K > 0 \\ \text{for } K < 0 \end{array} \quad (3.13b)$$

$$\begin{pmatrix} D \\ D' \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \begin{array}{l} \text{for } K = 0 \\ \text{for } K = 0 \end{array} \quad (3.13c)$$

To get the transformation matrix of the complete magnet one has to substitute the variable  $s$  by the length  $\ell$  of the magnet.

b) Drift space

$$\frac{1}{\rho} = 0, k = 0$$

$$M_x = M_z = \begin{pmatrix} 1 & \ell & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (3.14)$$

c) Quadrupole Magnet

The dispersion (3.13) vanishes since  $\frac{1}{\rho} = 0$ . With  $\varphi = \ell\sqrt{|k|}$  the transformation matrices are for  $k > 0$

$$M_x = \begin{pmatrix} \cosh \varphi & \frac{1}{\sqrt{|k|}} \sinh \varphi & 0 \\ \sqrt{|k|} \sinh \varphi & \cosh \varphi & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (3.15)$$

$$M_z = \begin{pmatrix} \cos \varphi & \frac{1}{\sqrt{|k|}} \sin \varphi & 0 \\ -\sqrt{|k|} \sin \varphi & \cos \varphi & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

These matrices describe horizontal defocusing, vertical focusing.

For  $k < 0$ , the matrices  $M_x$  and  $M_z$  are interchanged and we get horizontal focusing, vertical defocusing.

d) Thin-lens approximation

In many practical cases, the focal length  $f$  of the quadrupole magnet will be much larger than the length of the lens:

$$f = \frac{1}{k l} \gg l$$

Then the transfer matrices can be approximated by

$$M_x = \begin{pmatrix} 1 & 0 & 0 \\ \frac{1}{f} & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (3.16)$$

$$M_z = \begin{pmatrix} 1 & 0 & 0 \\ -\frac{1}{f} & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (3.17)$$

Note that these matrices describe a lens of zero length, i.e. they are derived from Eqs. (3.14) using  $\ell \rightarrow 0$  while keeping  $k \cdot \ell = \text{const.}$  The true length  $l$  of the lens has to be recovered by two

drift spaces  $\ell/2$  on either side, e.g.

$$M_z = \begin{pmatrix} 1 & \frac{\ell}{2} & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ -\frac{1}{f} & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & \frac{\ell}{2} & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 - \frac{\ell}{2f} & \ell - \frac{\ell^2}{4f} & 0 \\ -\frac{1}{f} & 1 - \frac{\ell}{2f} & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (3.18)$$

One might ask why the approximation has not been made by expanding  $\sin \varphi$ ,  $\cos \varphi$ , etc. in Taylor series and neglecting higher powers of  $\varphi$ . However terminating the Taylor series at some power results in a transfer matrix whose determinant is not unity. For instance, in third order we obtain

$$M_z = \begin{pmatrix} 1 - \frac{\ell}{2f} & \ell - \frac{\ell^2}{6f} & 0 \\ -\frac{1}{f} & 1 - \frac{\ell}{2f} & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

which **does not fulfil**  $\det M = 1$ . It will be shown later that this would violate Liouville's Theorem of phase-space conservation.

For accelerators in the TeV range, where  $1/\rho^2 \ll |k| \ll 1/\ell^2$ , the thin-lens approximation is excellent for the matrix description of the whole accelerator.

#### e) Dipole sector magnet

The matrices (3.12), (3.13) with  $k = 0$  describe a “hard edge” dipole, i.e. the magnet ends are perpendicular to the circular trajectory (Fig. 19).

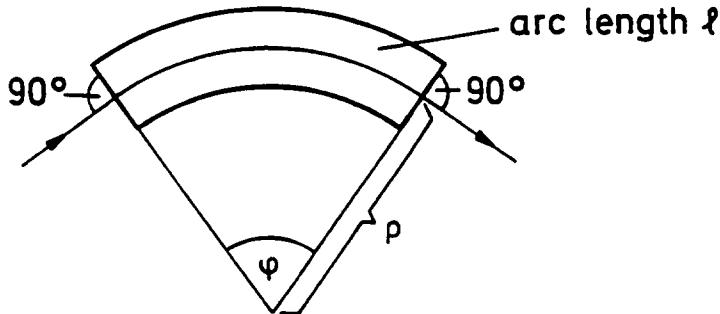


Figure 19: Dipole sector magnet

The transformation matrices are with  $\varphi = \frac{\ell}{\rho}$

$$M_x = \begin{pmatrix} \cos \varphi & \rho \sin \varphi & \rho(1 - \cos \varphi) \\ -\frac{1}{\rho} \sin \varphi & \cos \varphi & \sin \varphi \\ 0 & 0 & 1 \end{pmatrix} \quad M_z = \begin{pmatrix} 1 & \ell & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (3.19)$$

#### f) Rectangular dipole magnet

In practice, dipole magnets are often built straight with the magnet end plates not perpendicular to the central trajectory. A rectangular magnet can be derived from a sector magnet

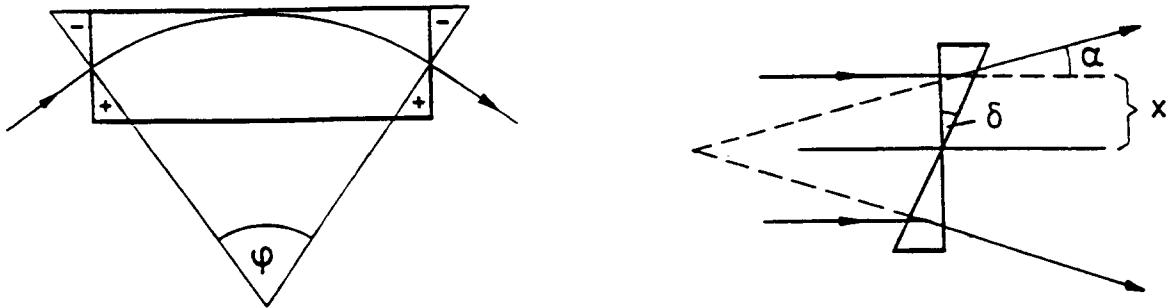


Figure 20: Rectangular dipole magnet and horizontally defocusing magnetic wedge

by superimposing at the entrance and exit a “magnetic wedge” of angle  $\delta = \varphi/2$ , as shown in Fig. 20.

The deflection angle in the magnetic wedge is

$$\alpha = \frac{\Delta\ell}{\rho} = \frac{x \tan \delta}{\rho} = \frac{x}{f}$$

It acts as a thin defocusing lens with  $1/f = (\tan \delta)/\rho$  in the horizontal plane, as a focusing length with the same strength in the vertical plane. The horizontal transformation matrix for a rectangular magnet is

$$M_x = \begin{pmatrix} 1 & 0 & 0 \\ \frac{1}{\rho} \tan \delta & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \cos \varphi & \rho \sin \varphi & \rho(1 - \cos \varphi) \\ -\frac{1}{\rho} \sin \varphi & \cos \varphi & \sin \varphi \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ \frac{1}{\rho} \tan \delta & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

For  $\varphi \ll 1$ ,  $\delta = \varphi/2$ :

$$M_x = \begin{pmatrix} 1 & \rho \sin \varphi & \rho(1 - \cos \varphi) \\ 0 & 1 & 2 \tan \varphi/2 \\ 0 & 0 & 1 \end{pmatrix} \quad M_z = \begin{pmatrix} \cos \varphi & \rho \sin \varphi & 0 \\ -\frac{1}{\rho} \sin \varphi & \cos \varphi & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (3.20)$$

Note that  $M_x$  is exact for  $\delta = \varphi/2$  while  $\varphi \ll 1$  has been used for  $M_z$  only. We conclude that in a rectangular magnet the weak horizontal focusing of a sector magnet is exactly compensated by the defocusing at the entrance and exit face. The magnet acquires, however, a weak vertical focusing of the same strength.

### g) Quadrupole doublet

The transformation matrix of a system of dipoles, quadrupoles and drift spaces is obtained by multiplying the matrices of each element in the correct order. An important example is a quadrupole doublet consisting of a focusing quadrupole, a drift space and a defocusing quadrupole. Figure 21 shows two trajectories (1,2) suggesting a tendency of both horizontal and vertical focusing in this kind of arrangement.

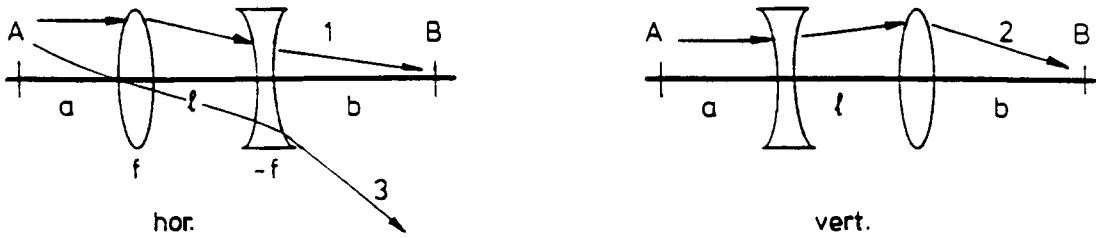


Figure 21: A quadrupole doublet consisting of a horizontally and a vertically focusing quadrupole magnet. Trajectories 1 and 2 suggest that there is a tendency of simultaneous focusing in both the horizontal and vertical directions.

The focusing action arises because trajectories entering parallel to the axis have a larger amplitude in the focusing than in the defocusing lens. Quadrupole doublets are indeed the simplest means of high energy beam focusing and imaging. We shall now derive the conditions for *simultaneous imaging in both* horizontal and vertical planes, treating the quadrupoles in the thin-lens approximation and assuming  $f_{foc} = -f_{defoc} = f$  for simplicity. The horizontal transfer matrix of the doublet is (for meaning of symbols see Fig. 21)

$$M_{doub,x} = \begin{pmatrix} 1 & 0 & 0 \\ \frac{1}{f} & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & \ell & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ -\frac{1}{f} & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 - \frac{\ell}{f} & \ell & 0 \\ -\frac{\ell}{f^2} & 1 + \frac{\ell}{f} & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (3.21)$$

The vertical transfer matrix is obtained if  $f$  is replaced by  $-f$ :

$$M_{doub,z} = \begin{pmatrix} 1 + \frac{\ell}{f} & \ell & 0 \\ -\frac{\ell}{f^2} & 1 - \frac{\ell}{f} & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (3.22)$$

The matrix element  $M_{21} = C' = -\ell/f^2$  is called the overall refractive power of the system and it is seen to be focusing in both planes. Somewhat sloppily one could say that a beam coming from infinity (i.e. all particles perfectly parallel to the  $s$ -axis,  $x'_0 = 0$ ) will be focused in both planes, as indicated by trajectories 1 and 2 in Fig. 21. The effective focal length  $f_{doub}$  for these particles is

$$f_{doub} = \frac{f^2}{\ell} \quad (3.23)$$

Trajectory 3 in Fig. 21, however, illustrates that there are trajectories as well which are not at all bent towards the beam axis. For practical applications one might therefore ask: What happens to particles emerging from a point  $A$  at a finite distance  $a$  from the first lens? Optical imaging requires that there is a point  $B$  at a distance  $b$  behind the second lens where all particles emerging from  $A$  will converge. The horizontal transfer matrix from  $A$  to  $B$  is

$$M_x = \begin{pmatrix} 1 & b & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} M_{\text{double},x} \begin{pmatrix} 1 & a & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 - \frac{\ell}{f} - \frac{\ell b}{f^2} & a + b + \ell + \frac{\ell b}{f} - \frac{\ell a}{f} - \frac{\ell ab}{f^2} & 0 \\ -\frac{\ell}{f^2} & 1 + \frac{\ell}{f} - \frac{\ell a}{f^2} & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (3.24)$$

Again,  $M_z$  is obtained if  $f$  is replaced by  $-f$ :

$$M_z = \begin{pmatrix} 1 + \frac{\ell}{f} - \frac{\ell b}{f^2} & a + b + \ell - \frac{\ell b}{f} + \frac{\ell a}{f} - \frac{\ell ab}{f^2} & 0 \\ -\frac{\ell}{f^2} & 1 - \frac{\ell}{f} - \frac{\ell a}{f^2} & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (3.25)$$

Imaging from  $A$  to  $B$  requires  $M_{12} = S \equiv 0$ . Because of  $\det M = 1$ , the matrix can then be written in the form

$$M = \begin{pmatrix} m & 0 & 0 \\ C' & \frac{1}{m} & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (3.26)$$

$m$  is called magnification of object  $A$  to image  $B$ . Obviously

$$x_B = m \cdot x_A \quad (\text{irrespective of } x'_A !) \quad (3.27)$$

Remark: If also  $M_{21} = C' = 0$  (i.e. zero overall refractive power), the system is called a telescopic system.

The condition  $S \equiv 0$  is satisfied if

$$\frac{b}{f} = \frac{\frac{a}{f} + \frac{\ell}{f} - \frac{\ell a}{f^2}}{\frac{\ell a}{f^2} - \frac{\ell}{f} - 1} \quad \text{for } M_x \quad (3.28)$$

and

$$\frac{a}{f} = \frac{\frac{b}{f} + \frac{\ell}{f} - \frac{\ell b}{f^2}}{\frac{\ell b}{f^2} - \frac{\ell}{f} - 1} \quad \text{for } M_z \quad (3.29)$$

There are two ways to interpret Eqs. (3.27-3.29):

- If the parameter set  $f, \ell, a, b$  is a solution in the horizontal plane, then the set  $f, \ell, b, a$  is a solution in the vertical plane, i.e. the rôles of  $a$  and  $b$  are interchanged. This means that in general horizontal and vertical images are in different planes. This is called “astigmatic” focusing.

Example: Consider  $\ell/f = 1$  and  $a/f = 3$ . Then  $b/f = 1$  and  $m_x = -1$  for imaging in the horizontal plane while  $b/f = 7/3$  and  $m_z = -1/3$  for imaging in the vertical plane. The vertical solution “conjugate” to the horizontal one would be  $\ell/f = 1$ ,  $a/f = 1$ ,  $b/f = 3$ , and  $m_z = -1$ . The latter one means, of course, that not only the horizontal and vertical images are in different distances from the doublet but also the respective objects.

- To get the horizontal and vertical images into the *same* plane (“stigmatic” focusing),  $a = b$  is required. Then

$$m_x = \frac{f + a}{f - a} = \frac{1}{m_z} \quad \text{stigmatic focusing} \quad (3.30)$$

Equal horizontal and vertical magnifications  $m_x = m_z$  are obviously impossible with stigmatic focusing, using quadrupole doublets. Note that, while  $f_{foc} = -f_{defoc}$  has been assumed for Eqs. (3.21-3.30), this latter statement applies for all kinds of quadrupole doublets.

Example:  $\ell/f = 4/3$ ,  $a/f = b/f = 2$  yields stigmatic focusing with  $m_x = -3$  and  $m_z = -1/3$ . Figure 22 illustrates particle trajectories in a stigmatic focusing quadrupole doublet.

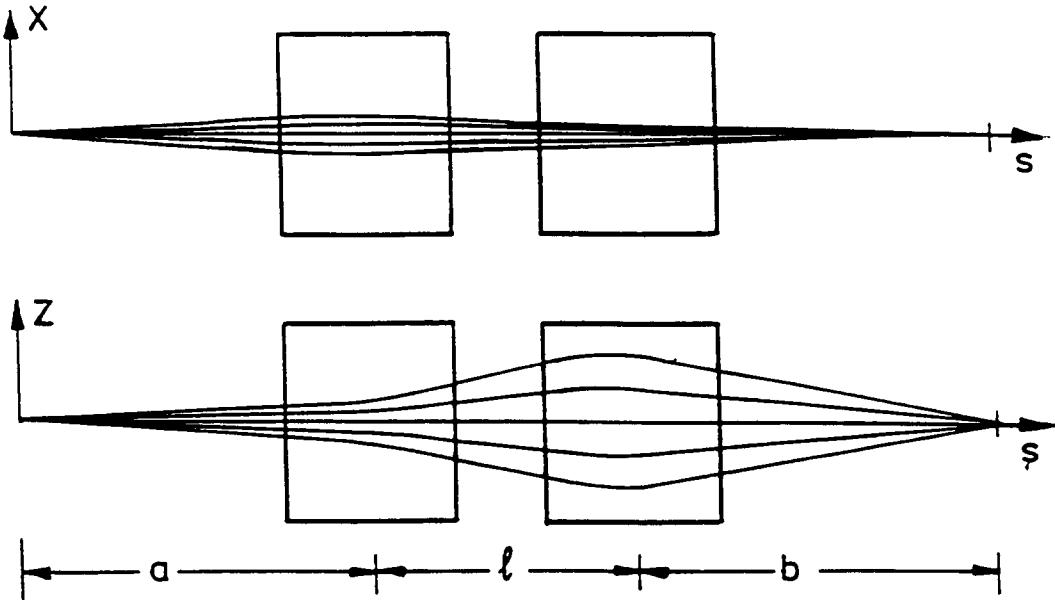


Figure 22: A stigmatic focusing quadrupole doublet showing particle trajectories that all start at the center of the object ( $x_A = z_A = 0$ ).

In circular accelerators and for beam transport along a transfer channel we are, in general, not interested in imaging at all. Instead, we require small (or at least finite) beam envelopes for any kind of particle source, i.e. for any location of the object plane. It will be shown in section 4.7 that, to achieve focusing in an alternating series of  $F$  and  $D$  quadrupoles, the separation between two quadrupoles must not be larger than twice the focal length

$$\ell < 2|f|$$

#### h) Accelerating section

Acceleration in the longitudinal direction is beyond the scope of this article. However, there is also an effect on the transverse motion, which is now briefly addressed.

Consider a section of length  $\ell$  with constant electric field  $E$ , in the longitudinal direction. The total momentum  $p$  of an ultrarelativistic particle ( $v \approx c = const$ ) then changes according to

$$\frac{dp}{ds} = \frac{e}{c} E_s = const \Rightarrow p(s) = p_0 + \frac{e}{c} E_s \cdot s$$

$p_0 = p(s = 0)$  is the momentum at the entrance of the section. The transverse motion of an ultrarelativistic particle is described by ( $y = x$  or  $z$ )

$$\frac{d}{ds} p_y = \frac{d}{ds} \left( \frac{p(s)}{c} v_y \right) = \frac{d}{ds} \left( p(s) \frac{dy}{ds} \right) = 0 \quad (3.31)$$

which is a different type of differential equation than (3.5), because  $p(s)$  is not constant any more. A first integral yields

$$p(s) \frac{dy}{ds} = \text{const} = y'_o p_o, \quad \text{thus} \quad \frac{dy}{ds} = y'(s) = \frac{y'_o p_o}{p_o + \frac{\epsilon}{c} E_s s} \quad (3.32)$$

Integrating once more we obtain

$$y(s) = y_o + y'_o \cdot \frac{cp_o}{eE_s} \ln \left( 1 + \frac{eE_s}{cp_o} s \right) \quad (3.33)$$

With Eqs. (3.32,3.33) and putting  $s = l$  the transfer matrix is

$$M_x = M_z = M = \begin{pmatrix} 1 & \frac{p_o}{\Delta p} \ell \cdot \ln \left( 1 + \frac{\Delta p}{p_o} \right) \\ 0 & \frac{p_o}{p_o + \Delta p} \end{pmatrix} \quad (3.34)$$

where  $\Delta p = \frac{\epsilon}{c} E_s \cdot l$  is the momentum gain in the accelerating section. The main complication with this transfer matrix is, that its determinant is not unity:

$$\det M = \frac{p_o}{p_o + \Delta p} \quad (3.35)$$

This is due to the fact that the equation of motion contains a first-derivative term, and it reflects the effect of adiabatic damping (see end of subsection 4.4). Instead of (3.34), one often uses a normalized matrix

$$M_x^* = M_z^* = M^* = \frac{M}{\sqrt{\det M}} = \frac{1}{\sqrt{\frac{p_o}{p_o + \Delta p}}} \begin{pmatrix} 1 & \frac{p_o}{\Delta p} \ell \cdot \ln \left( 1 + \frac{\Delta p}{p_o} \right) \\ 0 & \frac{p_o}{p_o + \Delta p} \end{pmatrix}$$

which has  $\det M^* = 1$ . This trick saves much of the formalism to be derived in the following sections, but if it is used one has to keep in mind, that in this case the *normalized* emittance is a conserved quantity and not the usual emittance, see Eq. (4.34). A similar complication arises if dispersion is included in the  $3 \times 3$  matrix version of (3.34).

## 4 BETATRON OSCILLATIONS

In this section, we shall first treat circular accelerators, because the formalism can be derived in a most stringent way for a periodic lattice. At the end (section 4.8) we will shed some light on the treatment of transfer lines or nonperiodic structures.

### 4.1 Stability criterion

We assume that the circular accelerator has a median plane, taken to be the horizontal plane  $z = 0$ , and that the magnetic guide and focusing fields are perpendicular to this plane. We further assume that there is a closed curve in this plane, called equilibrium orbit, on which a

particle with reference momentum  $p_0$  can move for an arbitrary number of revolutions. A point  $P$  on an arbitrary particle trajectory is characterized by the coordinates  $s, x$  and  $z$  defined by Fig. 2. The differential equations for  $x(s)$  and  $z(s)$  in linear approximation have been derived in section 3.

$$\begin{aligned} x'' - \left( k - \frac{1}{\rho^2} \right) x &= \frac{1}{\rho} \frac{\Delta p}{p_0} \\ z'' + kz &= 0 \end{aligned}$$

$k(s)$  and  $\rho(s)$  are periodic functions of  $s$  because the orbit is a closed curve. In the following we consider only particles of momentum  $p = p_0$ , i.e.  $\Delta p = 0$ .

Then both equations read

$$\begin{aligned} y'' + K(s)y &= 0 \\ K(s+L) &= K(s) \end{aligned} \tag{4.1}$$

$K(s)$  is a periodic function. The period  $L$  may be as large as the circumference  $C$  of the accelerator, but the accelerator may be composed of  $N$  identical sections or "cells" with  $C = N \cdot L$ . In that case  $K(s)$  has the cell length  $L$  as a period.

Remark: All large circular accelerators and storage rings possess indeed a periodic cell structure in the "regular" arcs. The straight sections usually deviate from the periodicity, for instance to accommodate interaction regions or wiggler magnets. These sections are matched to the arcs in such a way that the periodicity of the beta function (see below) is not disturbed.

The differential equation (4.1) is called Hill's equation. For the special case

$$K(s) = \text{const.} > 0$$

it describes a harmonic oscillation with two independent solutions

$$\cos(\sqrt{K}s) \text{ and } \sin(\sqrt{K}s).$$

In the general case  $K = K(s)$  we define the cosinelike and sinelike trajectories (3.7) and can express any solution  $y(s)$  of Eq. (4.1) in the form

$$y(s) = y_0 C(s) + y'_0 S(s)$$

We have already shown that  $y$  and  $y'$  at the point  $s$  can be obtained from  $y_0, y'_0$  by a matrix multiplication

$$\begin{pmatrix} y \\ y' \end{pmatrix}_s = M(s/s_0) \begin{pmatrix} y \\ y' \end{pmatrix}_{s_0} \tag{4.2}$$

with

$$M(s/s_0) = \begin{pmatrix} C(s) & S(s) \\ C'(s) & S'(s) \end{pmatrix}$$

and

$$\begin{pmatrix} C(s_0) & S(s_0) \\ C'(s_0) & S'(s_0) \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

The transfer matrix for two successive sections of magnets and drift spaces can be computed by matrix multiplication

$$M(s_2/s_0) = M(s_2/s_1) \cdot M(s_1/s_0) \quad (4.3)$$

Proof of Eq. (4.3):

call  $C_o(s)$  and  $S_o(s)$  the principal trajectories with reference point  $s_0$

$$i.e. \quad \begin{pmatrix} C_o(s_0) & S_o(s_0) \\ C'_o(s_0) & S'_o(s_0) \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Then

$$M(s_1/s_0) = \begin{pmatrix} C_o(s_1) & S_o(s_1) \\ C'_o(s_1) & S'_o(s_1) \end{pmatrix}$$

For arbitrary  $s_1 \neq s_0$ , the principal trajectories  $C_1(s), S_1(s)$  with reference point  $s_1$  are in general different from  $C_o(s), S_o(s)$  (see Fig. 23)

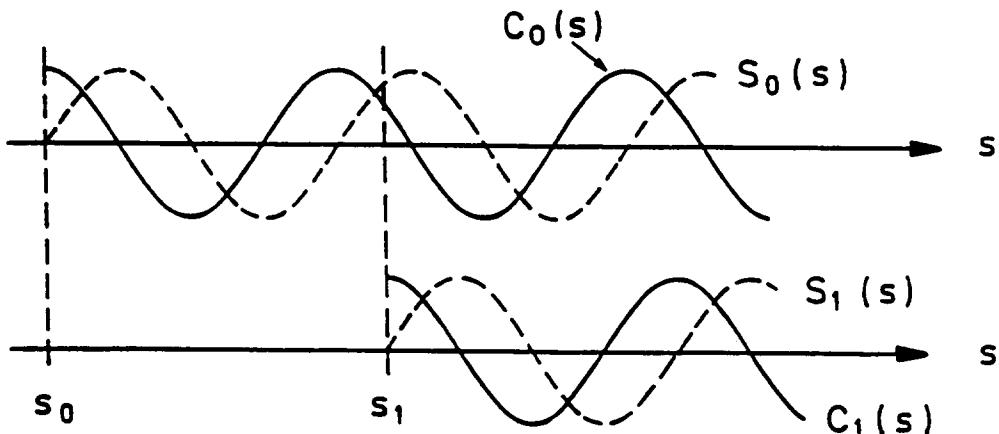


Figure 23: Cosine- and sinelike trajectories relative to different reference points  $s_0$  and  $s_1$

Since  $C_1(s)$  and  $S_1(s)$  are two linearly independent solutions, we can write

$$\begin{aligned} C_o(s) &= C_o(s_1) C_1(s) + C'_o(s_1) S_1(s) \\ S_o(s) &= S_o(s_1) C_1(s) + S'_o(s_1) S_1(s) \end{aligned}$$

Inserting  $s = s_2$  we obtain Eq. (4.3).

Of particular interest is the transfer matrix for a period  $L$  of the accelerator structure. We write it as

$$\mathbf{M}(s) \equiv M(s + L/s) \quad (4.4)$$

The transfer matrix for a full revolution is

$$M(s + NL/s) = (\mathbf{M}(s))^N$$

and for  $n$  turns

$$(\mathbf{M}(s))^{n \cdot N}$$

A necessary and sufficient condition for stable motion is that the elements of the matrix  $\mathbf{M}^{n \cdot N}$  remain bounded for  $n \rightarrow \infty$ . To derive a condition for this we consider the eigenvalues of the matrix  $\mathbf{M}$ .

$$\mathbf{M}Y = \lambda Y \quad Y = \begin{pmatrix} y \\ y' \end{pmatrix}$$

The eigenvalues are the solutions of the determinant equation

$$\det(\mathbf{M} - \lambda I) = 0 \quad (4.5)$$

Let us write  $\mathbf{M} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ , then Eq. (4.5) reads

$$\lambda^2 - \lambda(a + d) + (ad - bc) = 0 \quad (4.5a)$$

Now  $\det \mathbf{M} = ad - bc = 1$  (see section 3.2). Without loss of generality we may write

$$\cos \mu = \frac{1}{2} \text{trace } \mathbf{M} = \frac{1}{2}(a + d) \quad (4.6)$$

The new quantity  $\mu$  is real, if  $\frac{1}{2}|a + d| < 1$  and complex or imaginary for  $\frac{1}{2}|a + d| > 1$ . The two solutions of (4.5a) are

$$\lambda_1 = \cos \mu + i \sin \mu = e^{i\mu}, \quad \lambda_2 = \cos \mu - i \sin \mu = e^{-i\mu} \quad (4.7)$$

Let us now assume that  $|a + d| \neq 2$ ; then  $\cos \mu \neq 1$  and  $\sin \mu \neq 0$ .

In that case the matrix  $\mathbf{M}$  can be expressed in a very useful form (Twiss-matrix):

$$\mathbf{M} = I \cos \mu + J \sin \mu \quad (4.8)$$

Here  $I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$  is the unit matrix and  $J = \begin{pmatrix} \alpha & \beta \\ -\gamma & -\alpha \end{pmatrix}$   
with

$$\alpha = \frac{a - d}{2 \sin \mu}$$

$$\beta = b / \sin \mu \quad (4.9)$$

$$\gamma = -c/\sin \mu$$

From  $\det M = 1$  we get the relation

$$\beta\gamma - \alpha^2 = 1 \quad (4.10)$$

The matrix  $J$  has determinant 1, trace 0 and moreover the property

$$J^2 = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} = -I \quad (4.11)$$

The combination  $I \cos \mu + J \sin \mu$  has properties similar to the complex number  $e^{i\mu} = 1 \cdot \cos \mu + i \cdot \sin \mu$ ; in particular for any  $\mu_1, \mu_2$

$$(I \cos \mu_1 + J \sin \mu_1)(I \cos \mu_2 + J \sin \mu_2) = I \cos(\mu_1 + \mu_2) + J \sin(\mu_1 + \mu_2)$$

The  $n$ -th power of  $M$  is

$$M^n = I \cos(n\mu) + J \sin(n\mu) \quad (4.12)$$

From this relation we see immediately that the matrix elements of  $M^n$  for  $n \rightarrow \infty$  remain bounded if and only if  $\mu$  is real

Stability  $\iff |\text{trace } M| < 2 \iff \mu \text{ real}$

(4.13)

For real  $\mu$ , the matrix elements of  $M^n$  oscillate but remain bounded for any  $n$ . Furthermore,  $\mu = \text{real}$  guarantees that  $\alpha, \beta, \gamma$  are also real quantities. For complex or imaginary  $\mu$ , on the other hand,  $\cos n\mu$  and  $\sin n\mu$  increase exponentially and the elements of  $M^n$  do the same so that the motion becomes unbounded.

It should be noted that the trace of  $M$  and therefore the parameter  $\mu$  are independent of the reference point  $s$ . This can be seen as follows:

$$M(s_2 + L/s_1) = \underbrace{M(s_2 + L/s_2)}_{M(s_2)} M(s_2/s_1)$$

On the other hand

$$M(s_2 + L/s_1) = \underbrace{M(s_2 + L/s_1 + L)}_{M(s_2/s_1)} \underbrace{M(s_1 + L/s_1)}_{M(s_1)}$$

So  $M(s_2)$  differs from  $M(s_1)$  by a similarity transformation

$$M(s_2) = M(s_2/s_1) M(s_1) (M(s_2/s_1))^{-1} \quad (4.14)$$

Since the trace is not changed in such a transformation we have

$$\mu(s_2) = \mu(s_1)$$

i.e.  $\mu$  is independent of  $s$ .

The transfer matrix  $\mathbf{M}(s)$  as a whole does depend on the reference point  $s$ . In particular the elements of the matrix  $J$ :  $\alpha$ ,  $\beta$  and  $\gamma$  are periodic functions of  $s$  with period  $L$ .

$$\boxed{\begin{aligned}\mathbf{M}(s) &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \cos \mu + \begin{pmatrix} \alpha(s) & \beta(s) \\ -\gamma(s) & -\alpha(s) \end{pmatrix} \sin \mu = \begin{pmatrix} \cos \mu + \alpha \sin \mu & \beta \sin \mu \\ -\gamma \sin \mu & \cos \mu - \alpha \sin \mu \end{pmatrix} \\ \alpha(s+L) &= \alpha(s), \quad \beta(s+L) = \beta(s), \quad \gamma(s+L) = \gamma(s) \\ \mu &\text{ is independent of } s.\end{aligned}\tag{4.15}}$$

We have now succeeded in replacing the four matrix elements  $a, b, c, d$  fulfilling  $\det \mathbf{M} = ad - bc = 1$  by four parameters  $\alpha, \beta, \gamma, \mu$  satisfying  $\beta\gamma - \alpha^2 = 1$ . After the transfer matrix of the whole ring has been obtained by multiplication of individual matrices as described in section 3, the quantities  $\alpha, \beta, \gamma, \mu$  are easily calculated by Eqs. (4.6) and (4.9) at any reference point  $s$ . But what do we win? The advantage is that these parameters allow a simple description of the motion in terms of harmonic-oscillator language. This will be shown in the following sections.

## 4.2 Solution of the Hill equation

### Floquet Theorem

In section 3 we have solved the homogeneous equation of motion in a piecewise manner (with  $K(s) = \text{const}$  in each piece) and have combined them using matrix algebra. We shall now present a closed form solution of the transverse oscillations. From this we will gain an improved understanding of betatron oscillations. The equation

$$y'' + K(s)y = 0 \quad \text{with } K(s+L) = K(s)$$

has two independent solutions

$$y_1(s) = e^{i\mu s/L} p_1(s), \quad y_2(s) = e^{-i\mu s/L} p_2(s)\tag{4.16}$$

$\mu$  is called the characteristic coefficient of the differential equation and is given by

$$\cos \mu = \frac{1}{2} \text{trace } \mathbf{M}(s)$$

$p_1(s)$  and  $p_2(s)$  are periodic functions of  $s$

$$p_i(s+L) = p_i(s), \quad i = 1, 2$$

The solutions are bounded if  $\mu$  is real, i.e. if the trace of the transfer matrix over the magnet period of length  $L$  is less than 2.

The proof of Floquet's theorem goes in several steps.

a) We show that two special solutions  $y_1, y_2$  exist with the property

$$y_i(s+L) = m_i \cdot y_i(s) \quad m_i = \text{const.}$$

Proof: Take two arbitrary independent solutions  $g_1(s), g_2(s)$  and consider the new functions

$$h_i(s) = g_i(s + L), \quad i = 1, 2.$$

Because of the periodicity of  $K(s)$  the  $h_i(s)$  are also solutions. Therefore a nonsingular matrix  $A$  exists with

$$\begin{pmatrix} h_1 \\ h_2 \end{pmatrix}_s = \begin{pmatrix} g_1 \\ g_2 \end{pmatrix}_{s+L} = A \begin{pmatrix} g_1 \\ g_2 \end{pmatrix}_s$$

One can find a similarity transformation  $C$  which diagonalizes  $A$

$$CAC^{-1} \equiv B = \begin{pmatrix} m_1 & 0 \\ 0 & m_2 \end{pmatrix}$$

and define

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix}_s = C \begin{pmatrix} g_1 \\ g_2 \end{pmatrix}_s$$

Then

$$\begin{aligned} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}_{s+L} &= C \begin{pmatrix} g_1 \\ g_2 \end{pmatrix}_{s+L} = CA \begin{pmatrix} g_1 \\ g_2 \end{pmatrix}_s \\ &= CAC^{-1} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}_s = \begin{pmatrix} m_1 y_1 \\ m_2 y_2 \end{pmatrix}_s \end{aligned}$$

b) Now we establish a relation between  $A$  and the transfer matrix  $M(s) = M(s + L/s)$ . From

$$\begin{pmatrix} g_1 \\ g_2 \end{pmatrix}_{s+L} = A \begin{pmatrix} g_1 \\ g_2 \end{pmatrix}_s \quad \text{we get} \quad \begin{pmatrix} g'_1 \\ g'_2 \end{pmatrix}_{s+L} = A \begin{pmatrix} g'_1 \\ g'_2 \end{pmatrix}_s$$

Both equations can be combined in a single matrix equation

$$\begin{pmatrix} g_1 & g_2 \\ g'_1 & g'_2 \end{pmatrix}_{s+L} = \begin{pmatrix} g_1 & g_2 \\ g'_1 & g'_2 \end{pmatrix}_s \cdot A^T$$

$A^T$  is the transposed matrix of  $A$ .

From the definition of the transfer matrix we have on the other hand

$$\begin{pmatrix} g_1 & g_2 \\ g'_1 & g'_2 \end{pmatrix}_{s+L} = \underbrace{M(s + L/s)}_{M(s)} \begin{pmatrix} g_1 & g_2 \\ g'_1 & g'_2 \end{pmatrix}_s$$

So

$$A^T = G^{-1} \mathbf{M}(s) G \quad \text{with} \quad G = \begin{pmatrix} g_1 & g_2 \\ g'_1 & g'_2 \end{pmatrix},$$

Since the matrices  $A^T$  and  $\mathbf{M}(s)$  are connected by a similarity transformation they have the same eigenvalues:  $\lambda_1 = m_1, \lambda_2 = m_2$ .

So

$$y_i(s + L) = \lambda_i y_i(s) \quad i = 1, 2.$$

Now put

$$\lambda_1 = e^{i\mu}$$

( $\mu$  may be real or complex).

From  $\det \mathbf{M} = \lambda_1 \cdot \lambda_2 = 1$  follows

$$\begin{aligned} \lambda_2 &= e^{-i\mu} \quad \text{and} \quad \cos \mu = \frac{1}{2}(\lambda_1 + \lambda_2) \\ &= \frac{1}{2} \operatorname{trace} \mathbf{M} \end{aligned}$$

Finally we make the ansatz

$$y_1(s) = e^{+i\mu s/L} p_1(s), \quad y_2(s) = e^{-i\mu s/L} p_2(s)$$

It is then easy to see that the functions  $p_1(s), p_2(s)$  are periodic:  $p_i(s + L) = p_i(s)$ . This completes the proof of Floquet's theorem.

### 4.3 The beta function

The elements of the transfer matrix

$$\mathbf{M}(s) \equiv M(s + L/s) = \begin{pmatrix} m_{11}(s) & m_{12}(s) \\ m_{21}(s) & m_{22}(s) \end{pmatrix}$$

are clearly periodic functions of  $s$

$$m_{ik}(s + L) = m_{ik}(s)$$

The trace of  $\mathbf{M}$  and therefore the characteristic coefficient  $\mu$  of Hill's equation are independent of  $s$

$$\cos \mu = \frac{1}{2} \operatorname{trace} \mathbf{M} = \frac{1}{2}(m_{11} + m_{22})$$

If we write  $\mathbf{M}(s)$  in the Twiss form we see that the  $s$ -dependence is in the matrix  $J$ :

$$M(s) = I \cos \mu + \begin{pmatrix} \alpha(s) & \beta(s) \\ -\gamma(s) & -\alpha(s) \end{pmatrix} \sin \mu$$

We want to show that  $\alpha(s)$  and  $\gamma(s)$  can be related to the “beta function”  $\beta(s)$ . From the periodicity of  $M(s)$ :

$$\beta(s+L) = \beta(s) \quad (4.17)$$

The condition  $\det M(s) = 1$  allows to eliminate  $\gamma(s)$  (see Eq. (4.10)):

$$\boxed{\gamma(s) = \frac{1+\alpha^2(s)}{\beta(s)}} \quad (4.18)$$

In order to express  $\alpha(s)$  through the beta function (actually its derivative) we first derive first-order differential equations for the two eigenfunctions  $y_1, y_2$  of the Hill equation. From

$$\begin{pmatrix} y \\ y' \end{pmatrix}_{s+L} = M(s) \begin{pmatrix} y \\ y' \end{pmatrix}_s = e^{\pm i\mu} \begin{pmatrix} y \\ y' \end{pmatrix}_s$$

we obtain

$$y(s) \cos \mu + (y(s)\alpha + y'(s)\beta) \sin \mu = y(s) (\cos \mu \pm i \sin \mu)$$

$$y\alpha + y'\beta = \pm iy$$

$$\frac{y'}{y} = \frac{\pm i - \alpha}{\beta} \quad (4.19)$$

This last relation is valid with the positive sign for  $y_1(s)$  (Eq. (4.16)) and with negative sign for  $y_2(s)$ .

Let us take the equation of  $y_1$  and differentiate it logarithmically (we write  $y$  instead of  $y_1$ ).

$$\frac{y''}{y'} - \frac{y'}{y} = \frac{-\alpha'}{i-\alpha} - \frac{\beta'}{\beta}$$

$$y'' = -Ky \quad (\text{Hill equation})$$

$$\frac{y''}{y'} = -K \frac{y}{y'} = -\frac{K\beta}{i-\alpha} \quad (\text{from (4.19)}).$$

$$\frac{y''}{y'} - \frac{y'}{y} = -\frac{K\beta}{i-\alpha} - \frac{i-\alpha}{\beta} = \frac{-\alpha'}{i-\alpha} - \frac{\beta'}{\beta}$$

$$-K\beta^2 - (i-\alpha)^2 = -\alpha'\beta - \beta'(i-\alpha)$$

$$(\alpha^2 + K\beta^2 + \alpha\beta' - \alpha'\beta - 1) - i(2\alpha + \beta') = 0$$

The elements of  $\mathbf{M}(s)$  are real. Therefore we get

$$\alpha(s) = -\frac{1}{2}\beta'(s) \quad (4.20)$$

and

$$\alpha^2 + K\beta^2 + \alpha\beta' - \alpha'\beta - 1 = 0.$$

Using (4.20) to eliminate  $\alpha$  and  $\alpha'$  from the last equation we obtain the following nonlinear differential equation for the beta function

$$\frac{1}{2}\beta\beta'' - \frac{1}{4}\beta'^2 + K\beta^2 = 1 \quad (4.21)$$

We have now succeeded in expressing the elements of the transfer matrix  $\mathbf{M}(s)$  in terms of a single function  $\beta(s)$ , its derivative  $\beta'(s)$ , and a phase parameter  $\mu$ . In the following we shall see that the solutions of Hill's equation can also be written in terms of the beta function. This illustrates the basic rôle of this function.

According to Floquet's theorem there are two linearly independent solutions  $y_1, y_2$  of Hill's equation which can be written as a phase factor  $e^{\pm i\mu s/L}$  times a periodic function  $p_1(s), p_2(s)$ . These solutions fulfil the first-order differential equation (4.19)

$$\frac{y'}{y} = \frac{\pm i + \frac{1}{2}\beta'}{\beta}$$

After integration we get

$$y_{1,2}(s) = a\sqrt{\beta(s)}e^{\pm i\Phi(s)} \quad (4.22)$$

with  $\Phi'(s) = \frac{1}{\beta(s)}$ ,  $a = \text{const.}$

The most general solution of the Hill equation is a pseudo-harmonic oscillation; amplitude and wavelength depend on the coordinate  $s$  and are both given in terms of the beta function:

$$\text{amplitude } \propto \sqrt{\beta(s)} ; \quad \lambda(s) = 2\pi\beta(s) \quad (4.23)$$

The phase function  $\Phi(s)$  is also computed from  $\beta(s)$

$$\Phi(s) = \int_{s_0}^s \frac{dt}{\beta(t)} \quad (4.24)$$

Now we can also compute the characteristic coefficient  $\mu$  of Hill's equation.

$$y_1(s+L) = a\sqrt{\beta(s+L)}e^{+i\Phi(s+L)}$$

$$\beta(s+L) = \beta(s) \quad (\text{see (4.17)})$$

$$\Phi(s+L) = \int_{s_0}^{s+L} \frac{dt}{\beta(t)} = \int_{s_0}^s \frac{dt}{\beta(t)} + \int_s^{s+L} \frac{dt}{\beta(t)}$$

So

$$y_1(s + L) = y_1(s) \cdot \exp \left( +i \int_s^{s+L} \frac{dt}{\beta(t)} \right)$$

Using (4.16) we find

$$\mu = \int_s^{s+L} \frac{dt}{\beta(t)} \quad (4.25)$$

$\mu$  is the phase advance per period (of length  $L$ ).

The  $Q$  value (often denotes also as  $\nu$ ) is defined as the number of betatron oscillations per revolution. If the accelerator has  $N$  periods,  $Q$  is given by

$$Q = \frac{N\mu}{2\pi} = \frac{1}{2\pi} \oint \frac{ds}{\beta(s)} \quad (4.26)$$

In the above derivations it has been assumed that  $\beta(s)$  vanishes nowhere. This is guaranteed if  $|\cos \mu| < 1$  because in that case  $\alpha, \beta, \gamma$  are real and therefore the equation  $\beta\gamma - \alpha^2 = 1$  requires  $\beta \neq 0$ .

Remark: Note that the symbol  $\oint$  does not represent a line integral as defined in the frame of vector analysis. In accelerator physics it is, instead, frequently used to symbolize an integral over the full circumference of the circular accelerator.

Equation (4.26) means that, as a rule of thumb, the  $Q$ -value is given by the mean bending radius divided by the mean beta function.

A particle trajectory is described by a real solution of Hill's equation:

$$y(s) = a\sqrt{\beta(s)} \cos(\Phi(s) - \delta) \quad (4.27)$$

From  $\Phi' = \frac{1}{\beta}$ , that is  $d\Phi = \frac{1}{\beta(s)} ds = \frac{2\pi}{\lambda(s)} ds$  we see again that the local wavelength of this quasi-harmonic wave is  $\lambda(s) = 2\pi\beta(s)$ . Since in high energy accelerators the betatron wavelength is typically of the order of several 10 m,  $\beta$  is of the same order of magnitude.

The derivative of  $y(s)$  is

$$y'(s) = -\frac{a}{\sqrt{\beta}} (\alpha \cos(\Phi(s) - \delta) + \sin(\Phi(s) - \delta)) \quad (4.28)$$

$$\alpha = -\frac{1}{2}\beta'$$

$\delta$  is an arbitrary constant phase.

Now we consider a family of trajectories with the same amplitude  $a$  but with different phases  $\delta$ . At a fixed value of  $s$  we plot the vector

$$Y(\delta) = (y(s), y'(s))$$

as a function of the phase angle  $\delta$  in the  $(y, y')$ -phase space (horizontal phase space  $(x, x')$ , vertical phase space  $(z, z')$ ).

The equations

$$\begin{aligned} y &= a\sqrt{\beta} \cos(\Phi - \delta) \\ y' &= -\frac{a}{\sqrt{\beta}} (\sin(\Phi - \delta) + \alpha \cos(\Phi - \delta)) \end{aligned} \quad (4.29)$$

are a parametric representation of an ellipse in the  $(y, y')$  plane. If  $\delta$  varies between 0 and  $2\pi$ , the point  $Y = (y, y')$  moves around this ellipse (Fig. 24).

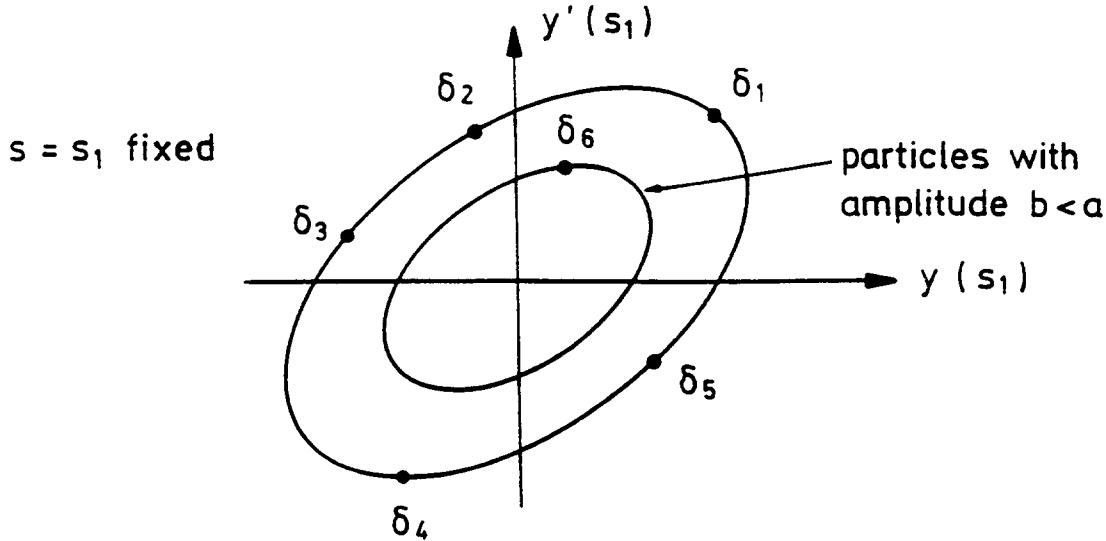


Figure 24: Phase-space ellipse for a family of particle trajectories with the same amplitude  $a$  but different phase angles  $\delta_1, \delta_2, \delta_3, \dots$  For trajectories with a smaller amplitude  $b$  one obtains a correspondingly smaller ellipse.

Alternatively we can consider only one particle and follow this particle for many successive revolutions. Then its phase-space coordinates at a fixed  $s = s_1$  also map out the phase-space ellipse, because for each turn in the accelerator the point  $(y, y')$  makes a noninteger number of  $Q$  turns in the phase plane (Fig. 25).

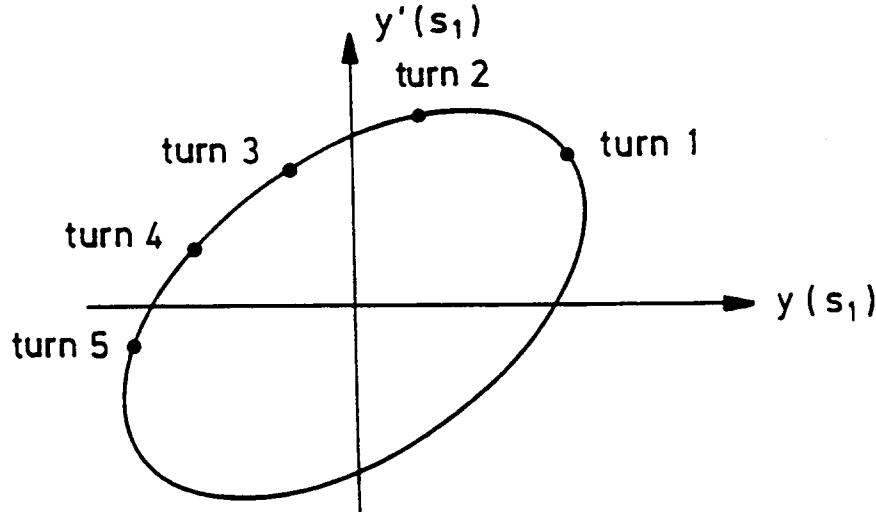


Figure 25: Phase-space ellipse for one particle after many turns in the accelerator. For each revolution of the particle, the point  $(y, y')$  makes  $Q$  turns in the phase plane (e.g. 6.2 turns in the case of the CERN PS).

#### 4.4 Emittance and beam envelope

The area of the phase-space ellipse is an important quantity. It is given by

$$\text{area} = \pi a^2 \quad (4.30)$$

We want to show that the area remains invariant when we transform the particle trajectories through the accelerator.

The ellipse equation is, in cartesian coordinate representation,

$$\gamma y^2 + 2\alpha y y' + \beta y'^2 = a^2 \quad (4.31)$$

This can be seen by inserting  $y, y'$  from Eq. (4.29) and using  $\beta\gamma - \alpha^2 = 1$ .

We want to show now that both sides of Eq. (4.31) remain constant when we transform  $y$  and  $y'$  through the accelerator. This constant is called the

Courant – Snyder invariant

$$\boxed{\gamma y^2 + 2\alpha y y' + \beta y'^2 = \text{const.}} \quad (4.32)$$

$$\text{or } \frac{1}{\beta} (y^2 + (\alpha y + \beta y')^2) = \text{const.}$$

Proof: let  $y(s)$  be an arbitrary real trajectory and  $y_1(s)$  the first eigenfunction (4.16) of Hill's equation. We form the Wronskian determinant

$$W = yy'_1 - y'y_1$$

$$\frac{dW}{ds} = W' = yy''_1 - y''y_1 = -K(yy_1 - yy_1) = 0$$

since both  $y, y_1$  fulfil Hill's equation. So  $W = \text{const.}$

$y_1$  fulfils the Eq. (4.19):  $y'_1 = \frac{i-\alpha}{\beta} y_1$ . Inserting this into the expression for  $W$  yields

$$W = y_1 \left( \frac{i-\alpha}{\beta} y - y' \right) = \text{const.}$$

$$WW^* = y_1 y_1^* \left( \frac{1}{\beta^2} y^2 + \left( \frac{\alpha}{\beta} y + y' \right)^2 \right) = \text{const.}$$

Now from Eq. (4.22):  $y_1 \sim \sqrt{\beta} e^{i\Phi}$ ,  $y_1 y_1^* \sim \beta$

$$\text{So we get } \frac{1}{\beta} (y^2 + (\alpha y + \beta y')^2) = \text{const.}$$

From the Courant-Snyder invariant we get the result that the area of the phase-space ellipse is invariant. The shape and orientation of the ellipse change when moving through the accelerator. In the center of a quadrupole of a regular FODO lattice the slope  $\beta'$  of the beta function is zero. Thus the main axes of the ellipse coincide with the  $y$  and  $y'$  axes (Fig. 26). (The term "FODO" lattice describes a periodic sequence of focusing and defocusing quadrupole magnets, with dipole magnets or drift spaces –denoted by "O"– between them, see Fig. 29.)

Let us now consider a particle beam centered around the reference orbit  $(y, y') = (0,0)$  and assume that the particle trajectories fill the  $(yy')$  phase space at a certain point  $s$  (e.g. the injection point) up to the phase ellipse with amplitude  $a$ . Then we know that all trajectories will remain inside or on the phase ellipse when the particles move around the accelerator. Since the area is invariant we get the important result that the phase-space density in the vicinity of any phase-space trajectory is invariant (Liouville's theorem). This is true under the condition that the particle energy is kept constant and that stochastic effects like beam-gas scattering or synchrotron radiation are neglected.

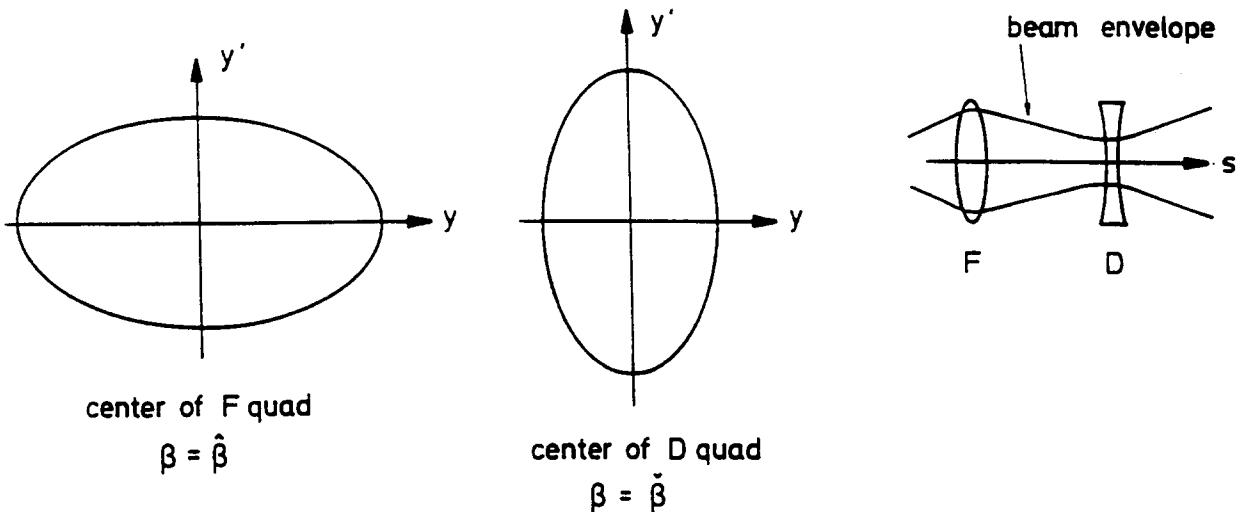


Figure 26: Phase-space ellipse in the center of a focusing or defocusing quadrupole

Since the area of the ellipse enclosing the beam in phase space is such an important quantity one introduces a new notion, the emittance  $\varepsilon$

$$\text{area of ellipse} = \pi \cdot \varepsilon \quad (4.33)$$

When the particles are accelerated, the emittance decreases inversely proportional to the momentum. This can be understood intuitively from the observation that only the longitudinal component of the momentum vector is increased in the accelerating cavities whereas the transverse components remain invariant, so that the beam divergence shrinks. This phenomenon is often called *adiabatic damping* which is somewhat misleading since no dissipative effect is involved. The energy dependence of the emittance can be derived in a rigorous way within the Hamiltonian formalism. The key point is that the canonically conjugate momenta of the position variables  $x$  and  $z$  are *not* the slopes  $x' = dx/ds$ ,  $z' = dz/ds$  but rather the transverse momenta  $p_x$ ,  $p_z$ . The phase-space trajectory in the  $(x, p_x)$  plane, corresponding to the emittance ellipse, is of course also an ellipse since  $p_x = p_0 x'$ . The area is  $\pi \cdot \varepsilon_N \cdot (m_0 c)$ . Here we have defined the *normalized emittance* by

$$\varepsilon_N = \left( \frac{p_0}{m_0 c} \right) \varepsilon \quad (4.34)$$

For proton and heavy-ion beams, and for electrons in linacs, the normalized emittance is the quantity that stays constant during acceleration. It remains even invariant when the beam is passed from an ion source through a whole chain of pre-accelerators into a large storage ring, provided the transitions are properly matched (see section 4.8).

In electron-positron storage rings, the (stochastic) emission of synchrotron radiation and the subsequent re-acceleration are the dominant factors determining the beam dimensions. In an ideal flat ring with complete decoupling of horizontal and vertical motion, the vertical emittance tends to zero while the horizontal emittance grows with the square of the energy. This is treated in more detail in the lectures by R. Walker, these proceedings.

Using the emittance  $\varepsilon$  we can write for a particle trajectory on the ellipse that encloses the beam

$$y(s) = \sqrt{\varepsilon\beta(s)} \cos(\Phi(s) - \delta) \quad (4.35)$$

The beam envelope is

$$E(s) = y_{max}(s) = \sqrt{\varepsilon} \sqrt{\beta(s)} \quad (4.36)$$

The beam divergence is

$$A(s) = y'_{max}(s) = \sqrt{\varepsilon} \sqrt{\frac{1 + \alpha^2(s)}{\beta(s)}} = \sqrt{\varepsilon} \sqrt{\gamma(s)} \quad (4.37)$$

These quantities are shown in Fig. 27.

Wherever  $\alpha = 0$ , the beam envelope  $E(s)$  has a local minimum ("waist") or maximum:  $\frac{d}{ds} E(s) = -\sqrt{\varepsilon}\alpha/\sqrt{\beta} = 0$ . At such a position, the beam divergence is  $A(s) = \sqrt{\varepsilon/\beta}$ . This offers another simple interpretation of the beta function at a beam waist: The beta function is just the ratio of beam size to beam divergence:

$$\beta(\text{beam waist}) = \frac{E(s)}{A(s)} = \frac{\sqrt{\varepsilon \cdot \beta}}{\sqrt{\varepsilon/\beta}} \quad (4.38)$$

Since  $\beta$  is typically of the order of meters, a beam of a few mm diameter will typically have a divergence of a milliradian.

Finally the equation of the ellipse enclosing the beam is

$$\gamma y^2 + 2\alpha y y' + \beta y'^2 = \varepsilon \quad (4.39)$$

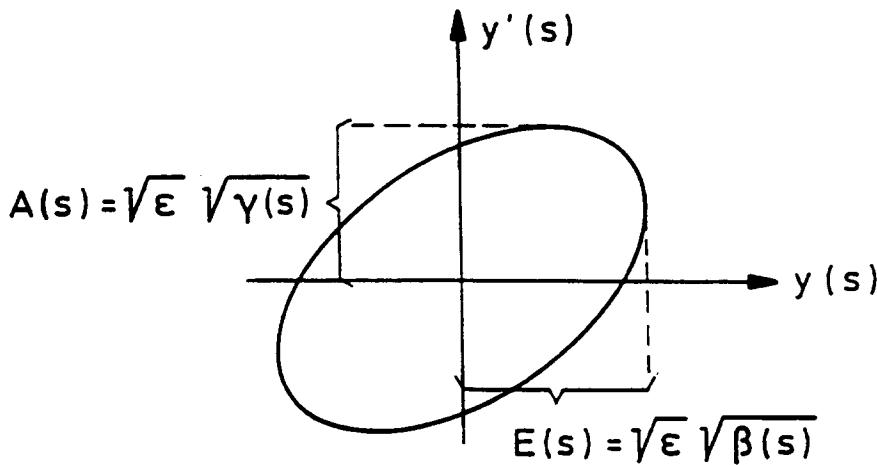


Figure 27: Beam envelope and divergence

**Remark:** There is a possible confusion here, since  $\varepsilon$  is often defined as the area rather than the (area)/ $\pi$ . To avoid such confusion, one often quotes numbers like  $(10\pi)\text{mm} \cdot \text{mrad}$ . With our definition of the emittance this implies  $\varepsilon = 10 \text{ mm} \cdot \text{mrad}$ . In the alternative definition one would have  $\varepsilon = 31.4 \text{ mm} \cdot \text{mrad}$ . Moreover, there are different conventions in the definition of the emittance for electron or proton beams, respectively. Traditionally, the electron emittance is defined for *one* standard deviation of the phase-space distribution function while for protons *two* standard deviations are used.

## 4.5 Transformation of $\alpha, \beta, \gamma$

The phase ellipse at a point  $s$  is characterized by the parameters  $\alpha(s)$ ,  $\beta(s)$ ,  $\gamma(s)$  and the emittance  $\varepsilon$ . It is important to know the beam envelope and divergence at each location of the accelerator. Therefore it is of great interest to know how to transform  $\alpha, \beta, \gamma$  through the optical system.

We evaluate the invariant  $\varepsilon$  at a reference point  $s_0$  and an arbitrary other point  $s$ .

$$\varepsilon = \gamma y^2 + 2\alpha y y' + \beta y'^2 = \gamma_0 y_0^2 + 2\alpha_0 y_0 y'_0 + \beta_0 y'^2_0$$

Using the principal trajectories we can relate  $y(s)$ ,  $y'(s)$  to  $y_0 = y(s_0)$ ,  $y'_0 = y'(s_0)$

$$\begin{pmatrix} y \\ y' \end{pmatrix} = \begin{pmatrix} C & S \\ C' & S' \end{pmatrix} \begin{pmatrix} y_0 \\ y'_0 \end{pmatrix}$$

The inverse transformation matrix is

$$\begin{pmatrix} S' & -S \\ -C' & C \end{pmatrix} \text{ because of } \begin{vmatrix} C & S \\ C' & S' \end{vmatrix} = 1.$$

So

$$y_0 = S'y - Sy'$$

$$y'_0 = -C'y + Cy'$$

$$\begin{aligned} \varepsilon = & \gamma_0 (S'y - Sy')^2 + 2\alpha_0 (S'y - Sy')(-C'y + Cy') \\ & + \beta_0 (-C'y + Cy')^2 = \gamma y^2 + 2\alpha y y' + \beta y'^2 \end{aligned}$$

Comparing the coefficients yields:

$$\begin{aligned} \beta(s) &= C^2 \beta_0 - 2SC\alpha_0 + S^2\gamma_0 \\ \alpha(s) &= -CC'\beta_0 + (SC' + S'C)\alpha_0 - SS'\gamma_0 \\ \gamma(s) &= C'^2 \beta_0 - 2S'C'\alpha_0 + S'^2\gamma_0 \end{aligned}$$

In matrix notation

$$\begin{pmatrix} \beta \\ \alpha \\ \gamma \end{pmatrix} = \begin{pmatrix} C^2 & -2SC & S^2 \\ -CC' & SC' + S'C & -SS' \\ C'^2 & -2S'C' & S'^2 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \alpha_0 \\ \gamma_0 \end{pmatrix} \quad (4.40)$$

By means of this matrix equation one can transform the beta function piecewise through the magnets and drift spaces. It is useful to start at a symmetry point  $s_0$  of the machine where the derivative of  $\beta$  vanishes.

$$\begin{pmatrix} \beta_0 \\ \alpha_0 \\ \gamma_0 \end{pmatrix} = \begin{pmatrix} \beta_0 \\ 0 \\ \frac{1}{\beta_0} \end{pmatrix} \quad \alpha_0 = -\frac{1}{2}\beta'_0 = 0$$

In a drift space we get

$$\beta(s) = \beta_0 - 2\alpha_0 s + \gamma_0 s^2, \quad \alpha(s) = \alpha_0 - \gamma_0 s, \quad \gamma(s) = \gamma_0.$$

If now  $s_0$  is a symmetry point like the interaction point of a colliding-beam storage ring we have

$$\beta(s) = \beta_0 + \frac{s^2}{\beta_0} \quad (4.41)$$

In order to get high luminosity one needs a small beam size at the interaction point (IP) and thus  $\beta_0$  as small as possible. Due to Eq. (4.41) then  $\beta$  becomes large in the quadrupoles next to the IP. Therefore one needs large aperture quadrupoles, located as close to the IP as possible. This in turn means that the focusing strength must be large. The limiting factor in these so called "mini-beta schemes" turns out to be, in most cases, the large chromatic errors of the quadrupole magnets involved, see section 5.3 [9].

We have already seen that the beta function has the same periodicity as Hill's equation:

$$y'' + K(s)y = 0$$

$$K(s+L) = K(s), \quad \beta(s+L) = \beta(s)$$

The solutions (4.27), i.e. the particle trajectories, however, are in general not periodic (Fig. 28). On the contrary, an integer  $Q$  value, i.e. an integer number of betatron oscillations per revolution, has to be carefully avoided because otherwise a small perturbation at a certain point  $s = s_1$  would be seen by the particle at every passage with the same betatron phase angle. This would lead to a resonance-like increase in the betatron amplitude and finally to a loss of the particle.

## 4.6 Principal trajectories in terms of amplitude and phase function

We denote with  $C(s)$  and  $S(s)$  the cosinelike and sinelike trajectories relative to the start point  $s = s_0$

$$C(s_0) = 1 \quad S(s_0) = 0$$

$$C'(s_0) = 0 \quad S'(s_0) = 1$$

Both can be represented by the form (4.27) of a real trajectory

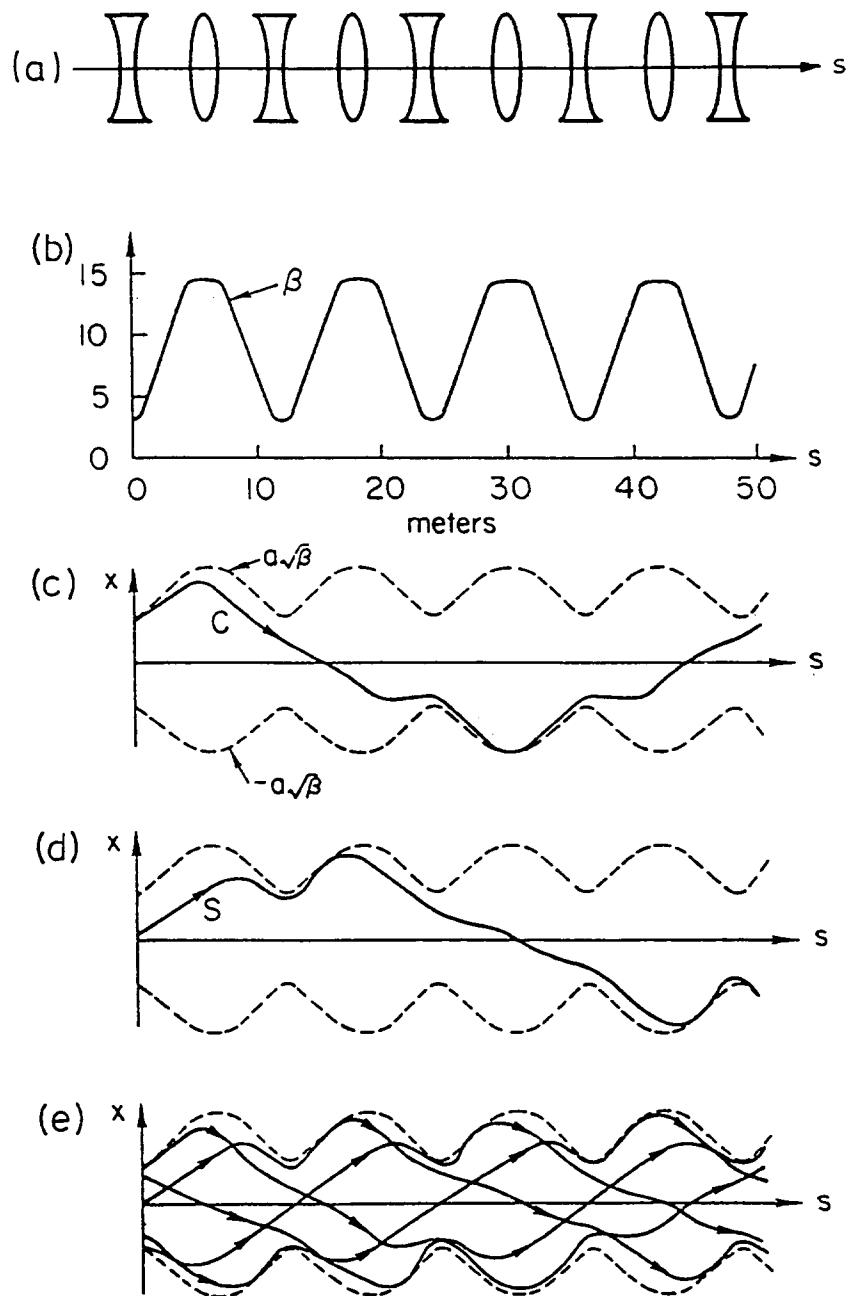


Figure 28: a) A regular FODO lattice of focusing and defocusing lenses b) Beta-function. c) Cosine-like trajectory for  $s = 0$ . d) Sine-like trajectory for  $s = 0$ . e) One trajectory on several successive revolutions.  
 (According to M. Sands, The Physics of Electron Storage Rings, SLAC-121).

$$C(s) = \sqrt{\beta(s)}(a \cos(\Phi(s) - \Phi_0) + b \sin(\Phi(s) - \Phi_0))$$

with  $\Phi_0 = \Phi(s_0)$

$$\begin{aligned} C' = & \frac{1}{2}\beta^{-\frac{1}{2}}\beta'(a \cos(\Phi - \Phi_0) + b \sin(\Phi - \Phi_0)) \\ & + \sqrt{\beta}(-a \sin(\Phi - \Phi_0) + b \cos(\Phi - \Phi_0))\Phi' \end{aligned}$$

From  $C(s_0) = 1$ ,  $C'(s_0) = 0$  we get

$$a = \frac{1}{\sqrt{\beta_0}}, \quad b = \alpha_0 a = \frac{\alpha_0}{\sqrt{\beta_0}}$$

$S(s)$  can be computed similarly. The transformation matrix becomes

$$\begin{pmatrix} C(s) & S(s) \\ C'(s) & S'(s) \end{pmatrix} = \begin{pmatrix} \sqrt{\frac{\beta}{\beta_0}}(\cos \Delta\Phi + \alpha_0 \sin \Delta\Phi) & \sqrt{\beta\beta_0} \sin \Delta\Phi \\ \frac{1}{\sqrt{\beta\beta_0}}((\alpha_0 - \alpha) \cos \Delta\Phi - (1 + \alpha\alpha_0) \sin \Delta\Phi) & \sqrt{\frac{\beta_0}{\beta}}(\cos \Delta\Phi - \alpha \sin \Delta\Phi) \end{pmatrix} \quad (4.42)$$

$$\text{with } \Delta\Phi = \Phi(s) - \Phi(s_0)$$

An interesting special case is the transformation matrix for one revolution. For simplicity we start at a symmetry point

$$s = s_0 : \quad \beta'_0 = 0, \quad \alpha_0 = 0$$

Then

$$\begin{pmatrix} C & S \\ C' & S' \end{pmatrix} = \begin{pmatrix} \cos 2\pi Q & \beta_0 \sin 2\pi Q \\ -\frac{1}{\beta_0} \sin 2\pi Q & \cos 2\pi Q \end{pmatrix} \quad (4.42a)$$

## 4.7 Transformation through a FODO cell

Most high energy accelerators or storage rings have a periodic sequence of quadrupole magnets of alternating polarity in the arcs, see Fig. 29. The dipole magnets are put in between. Here we want to discuss the optical properties of a “FODO” cell disregarding the bending magnets (they are treated as drift spaces). The more general case will be considered in section 6.

We work in the “thin-lens” approximation, see section 3.3d.

$$M_F = \begin{pmatrix} 1 & 0 \\ -\frac{1}{f} & 1 \end{pmatrix} \quad M_0 = \begin{pmatrix} 1 & L/2 \\ 0 & 1 \end{pmatrix} \quad M_D = \begin{pmatrix} 1 & 0 \\ +\frac{1}{f} & 1 \end{pmatrix}$$

The transformation matrix of the cell is

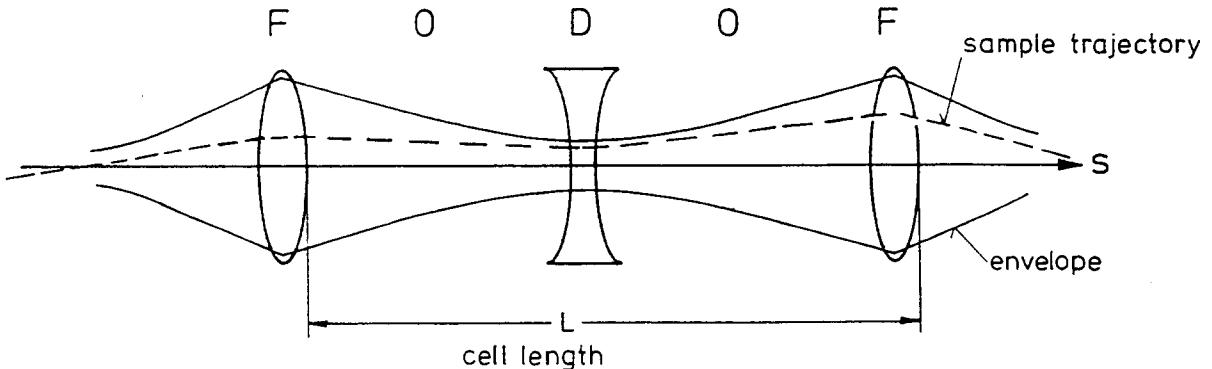


Figure 29: FODO cell

$$\mathbf{M} = M_F \cdot M_O \cdot M_D \cdot M_O$$

$$\mathbf{M} = \begin{pmatrix} 1 + \frac{L}{2f} & L + \frac{L^2}{4f} \\ -\frac{L}{2f^2} & 1 - \frac{L}{2f} - \frac{L^2}{4f^2} \end{pmatrix} \quad (4.43)$$

If we compare this to the Twiss representation (4.15) of the transformation matrix over one period we obtain

$$\cos \mu = \frac{1}{2} \text{trace } \mathbf{M} = 1 - \frac{L^2}{8f^2}$$

$$\cos \mu = 1 - 2 \sin^2 \frac{\mu}{2} \quad (4.44)$$

$$|\sin \frac{\mu}{2}| = \frac{L}{4f}$$

This equation allows one to compute the phase advance per cell from the cell length and the focal length of the quadrupoles.

Stability requires (see (4.13)):  $|\cos \mu| < 1$

i.e.  $L/4f < 1$  Stability:  $f > L/4$  (4.45)

For a phase advance of  $90^\circ$  per cell

$$f = \frac{1}{\sqrt{2}} \cdot \frac{L}{2}$$

Example: HERA proton ring

$$L = 47 \text{ m}, \quad k \simeq 0.032 \text{ m}^{-2}, \quad \ell_{quad} \simeq 1.9 \text{ m}$$

$$f \simeq 16.4 \text{ m} \quad \mu \simeq 90^\circ \quad \beta = \left( L + \frac{L^2}{4f} \right) / \sin \mu \simeq 80.7 \text{ m} \quad (\text{in focusing quadrupole})$$

The limiting case  $L = 4f$  of (4.45) has a simple interpretation. It is well known from optics that an object at a distance  $a = 2f$  from a focusing lens has its image at  $b = 2f$ . A regular arrangement of focusing lenses with gaps  $L = 4f$  in between provides thus a sequence of point-to-point imaging (Fig. 30). Defocusing lenses that are inserted at these image points have no

effect at all because they are traversed on the axis. If however the lens system is moved further apart ( $L > 4f$ ), this is no more true and the divergence of the light or particle beam is increased by every defocusing lens.

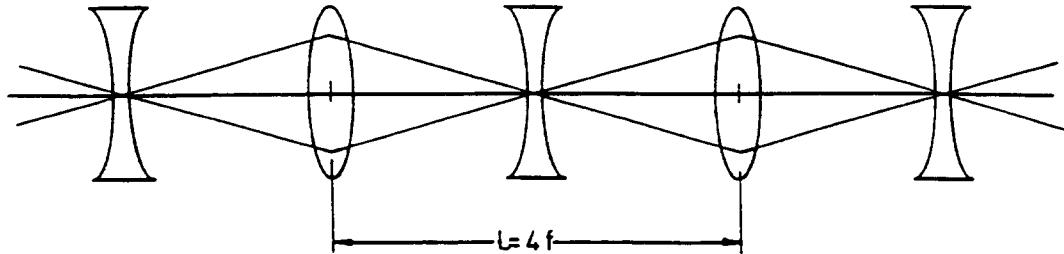


Figure 30: Point-to-point imaging in an arrangement of focusing lenses with distances  $L = 4f$ . The defocusing lenses have no effect if a point-like object is located exactly on the axis at distance  $2f$  from a focusing lens.

As another example of the application of FODO cells, Fig. 31 illustrates the evolution of the phase space ellipse in a very simple circular accelerator consisting of just one FODO cell.

#### 4.8 Non-periodic beam optics

In the previous sections, the Twiss parameters  $\alpha, \beta, \gamma, \mu$  have been derived for a periodic, circular accelerator. The condition of periodicity was essential for the definition of the beta function, see Eq. (4.17). Quite often, however, a particle beam moves only once along a beam transfer line, but one is nonetheless interested in quantities like beam envelopes and beam divergence. We will now show that the quantities  $\alpha, \beta, \gamma, \mu$  are useful in beam transfer lines as well. It will turn out that the main difference is that in transfer lines the beta function is no longer uniquely determined by the transfer matrix, but also depends on initial conditions which have to be specified in an adequate way.

There is a very simple Gedankenexperiment to show that the Twiss parameters are useful in transfer lines as well: consider a circular accelerator which contains the transfer line as a part of its lattice. Then it is obvious that the optics in the transfer section can be described in terms of Twiss parameters. Now one has to realize that the Twiss parameters in the transfer section will depend on the *complete* revolution matrix. Thus, since there is large arbitrariness in how the transfer line is complemented, the Twiss parameters in the transfer line are not at all well defined. On the other hand Eq. (4.40) shows, that the Twiss parameters are perfectly known in the whole transfer line if only  $\alpha$  and  $\beta$  are determined somehow at the entrance of the line. How can this be done reasonably?

Since  $\alpha$  and  $\beta$  are related to the beam divergence and beam envelope, a sensible approach is to derive the initial conditions of  $\alpha$  and  $\beta$  from the actual phase space distribution of the beam entering the transfer line. In Fig. 32 we have sketched such a distribution.

The lines of constant phase space density are often well fitted by properly parametrized ellipses. (An ellipse is indeed one of the most simple parametrizations one can think of; the main reason for choosing an ellipse, however, will readily be seen.)  $1/\pi$  times the area of the ellipse which contains a certain fraction of the particles (say 90 %) is called the beam emittance. The most general

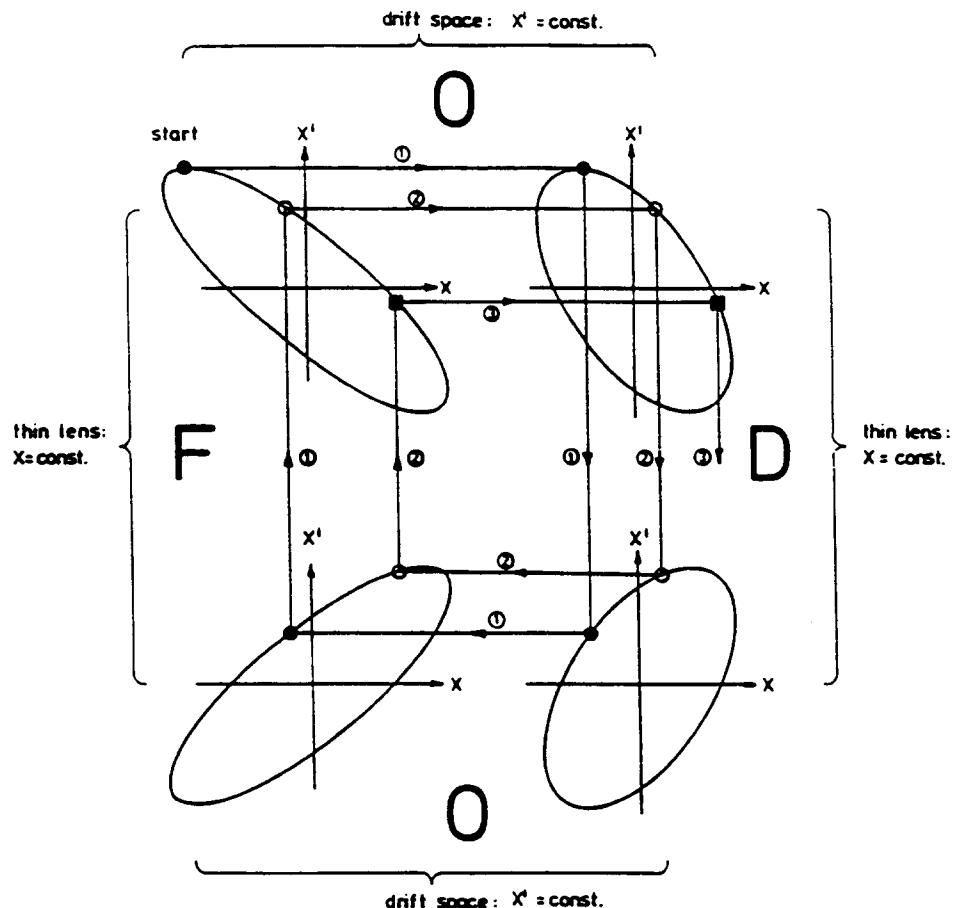


Figure 31: Phase-space dynamics in a simple circular accelerator consisting of one FODO cell. The two  $180^\circ$  bending magnets are assumed to be located in the drift spaces. Their weak focusing contribution is neglected. The periodicity of  $\alpha, \beta, \gamma$  is reflected by the fact that the phase-space ellipse is transformed into itself after each turn. An individual particle trajectory, however, which starts, for instance, somewhere on the ellipse at the exit of the focusing quadrupole (small circle), is seen to move on the ellipse from turn to turn as determined by the phase angle  $\mu$ . Thus, an individual particle trajectory is not periodic, while the envelope of a whole beam is.

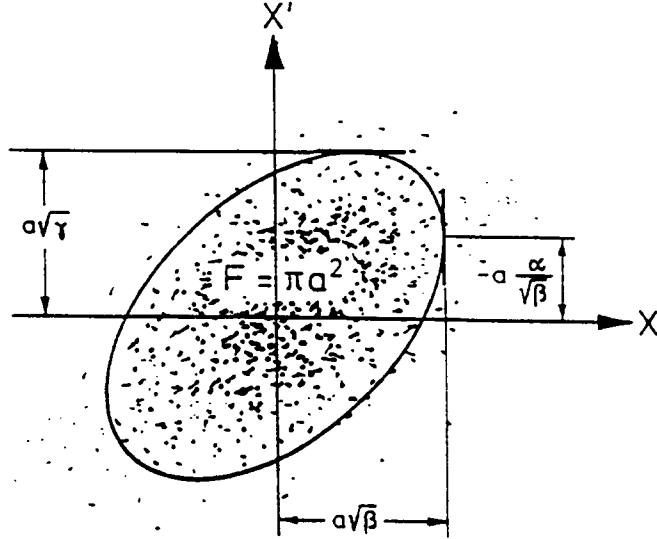


Figure 32: A particle beam is often reasonably well described by a two dimensional Gaussian distribution in phase space. The lines of constant phase-space density are then ellipses. Since the phase-space density decreases only slowly with amplitude, the phase-space area containing *all* particles might be hard to determine (experimentally as well as theoretically). Also, it is not the quantity relevant for most of the applications. Therefore, the emittance is defined as  $1/\pi$  times the phase-space area containing a certain fraction of the particles (e.g. 90 %).

ellipse parametrization in phase space is given by ( $y = x$  or  $z$ )

$$\gamma y^2 + 2\alpha yy' + \beta y'^2 = a^2 \quad (4.46)$$

Since only three free parameters are needed, we can impose another condition:

$$\beta\gamma - \alpha^2 = 1 \quad (4.47)$$

With this normalization of the parameters  $\alpha, \beta, \gamma$ , the parameter  $a$  has the simple meaning that  $\pi a^2$  is the ellipse area, thus

$$a^2 = \varepsilon = \text{emittance} \quad (4.48)$$

It is obvious from Eqs. (4.46-4.48) that our beam ellipse parameters  $\alpha, \beta, \gamma, a$  satisfy the same equations as the Twiss parameters defined for a circular accelerator, see Eqs. (4.18, 4.30, 4.31). In a circular accelerator, however,  $\alpha, \beta, \gamma$  are completely determined by the magnet optics and the condition of periodicity while beam properties are not at all involved (only  $\varepsilon$  is chosen to fit the actual beam size). In a transfer line optics, on the other hand, we can choose  $\beta, \beta', \varepsilon$  at the entrance to fit best the incoming beam.

From then on the optics calculation proceeds the same way as in the circular accelerator:

Linear optics performs a linear transformation of phase-space coordinates,

$$\begin{pmatrix} y \\ y' \end{pmatrix}_1 = M(s_1/s_0) \begin{pmatrix} y \\ y' \end{pmatrix}_0 = \begin{pmatrix} C & S \\ C' & S' \end{pmatrix} \begin{pmatrix} y \\ y' \end{pmatrix}_0 \quad (4.49)$$

The beam ellipse at the entrance  $s_0$  is therefore transformed into another ellipse at  $s_1$ :

$$\gamma_1 y_1^2 + 2\alpha_1 y_1 y'_1 + \beta_1 y'_1^2 = a_1^2 = \gamma_0 y_0^2 + 2\alpha_0 y_0 y'_0 + \beta_0 y'_0^2 = a_0^2 \quad (4.50)$$

Using Eq. (4.49) this can be written in the well known form (see Eq. (4.40))

$$\begin{pmatrix} \beta_1 \\ \alpha_1 \\ \gamma_1 \end{pmatrix} = \begin{pmatrix} C^2 & -2SC & S^2 \\ -CC' & SC' + S'C & -SS' \\ C'^2 & -2S'C' & S'^2 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \alpha_0 \\ \gamma_0 \end{pmatrix} \quad (4.51)$$

Using  $\det \mathbf{M} = 1$  it is easy to prove that

$$\beta_1 \gamma_1 - \alpha_1^2 = 1$$

i.e.  $\pi a_1^2$  still means the phase-space area. By virtue of Eq. (4.50)

$$a_0^2 = \varepsilon_0 = a_1^2 = \varepsilon_1$$

i.e. beam emittance is preserved.

Compared to the previous sections, there is nothing new with Eqs. (4.46-4.51). It is just the interpretation of the Twiss parameters which makes a difference.

## 5 MOTION OF PARTICLES WITH MOMENTUM DEVIATION

The central design orbit of a circular accelerator is a closed curve that goes through the center of all quadrupoles (assuming that the magnets are well aligned). This orbit is a possible particle trajectory: particles with nominal momentum  $p_0$  which start at some point with zero displacement and zero slope will move on the design orbit for an arbitrary number of revolutions. However, particles with  $p = p_0$  which start with non-vanishing initial conditions will conduct betatron oscillations about the orbit. Their path around the ring does not close onto itself since the Q-value is non-integer.

Now we consider a particle with a larger momentum  $p > p_0$ . The design orbit is no more a possible trajectory for this particle. This is easy to understand for a weakly focusing machine (or in a homogeneous field): particles with larger momentum need a circle of larger radius on which they can move indefinitely (see Fig. 33).

If they do not start on exactly that circle, they will perform betatron oscillations about this new, larger circle which therefore is the reference orbit for the particles with the given momentum deviation  $dp = p - p_0$ . Since its radial distance  $x(s, dp/p_0)$  from the design orbit is proportional to  $dp/p_0$ , it is practical to divide by the momentum deviation and define the closed dispersion orbit  $D(s)$  by the equation

$$x(s, \frac{dp}{p_0}) = D(s) \cdot \frac{dp}{p_0}$$

In a machine with weak focusing,  $D(s)$  is of course a constant. In strongly focusing machines such an orbit exists, too, but it looks a lot more complicated. The focusing quadrupoles do not permit the particles to deviate too much from the central orbit and bend the trajectory towards the central orbit whereas the defocusing quadrupoles bend it away. It is then easy to understand that the closed dispersion orbit has its maximum deviation from the design orbit in the center

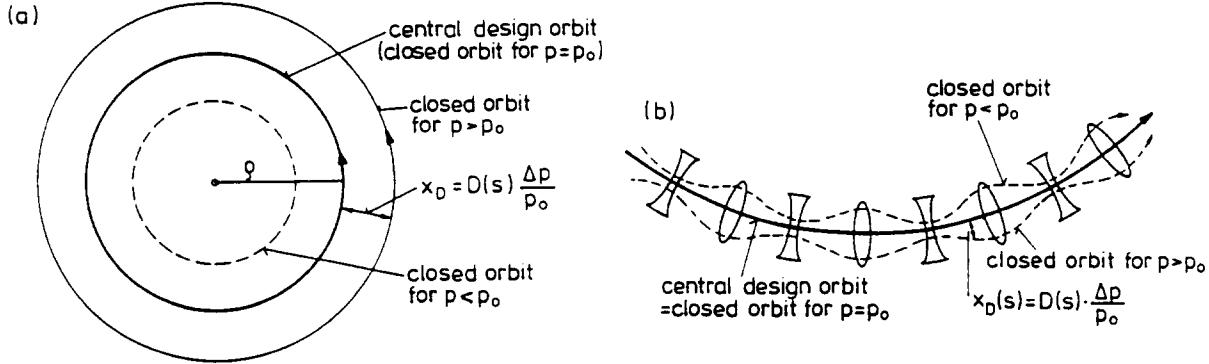


Figure 33: Closed orbit for particles with momentum  $p \neq p_0$  in a weakly (a) and strongly (b) focusing circular accelerator.

of the focusing quadrupoles and its minimum deviation in the defocusing quadrupoles. This is sketched in Fig. 33 b.

In the following we want to derive the mathematical expression for the closed dispersion orbit.

### 5.1 Closed orbit for $\Delta p \neq 0$

A particle with  $\Delta p = p - p_0 \neq 0$  satisfies the inhomogeneous Hill equation for the horizontal motion

$$x'' + K(s)x = \frac{1}{\rho} \frac{\Delta p}{p_0} \quad (5.1)$$

The total deviation of the particle from the reference orbit of the machine can be written as

$$x(s) = x_D(s) + x_\beta(s) \quad (5.2)$$

Here  $x_D(s) = D(s) \cdot \frac{\Delta p}{p_0}$  describes the deviation of the closed orbit for off-momentum particles with a fixed  $\Delta p$  from the reference orbit;  $x_\beta(s)$  describes the betatron oscillation around this closed dispersion orbit.

$D(s)$  is the “periodic dispersion”. It satisfies just the same differential equation as the dispersion trajectory in section 3.2:

$$D'' + K(s)D = \frac{1}{\rho(s)} \quad (5.3)$$

But now we impose *periodic* boundary conditions

$$D(s + C) = D(s), \quad D'(s + C) = D'(s) \quad (5.4)$$

$C = NL$  is the circumference of the machine.

The periodic dispersion is often denoted as  $\eta(s)$ .

In general, also the equation for the vertical coordinate  $z$  contains a nonvanishing right hand side, e.g. due to field and alignment errors in the magnets. We therefore consider more generally the equation

$$y'' + K(s)y = F(s) \quad (5.5)$$

$K(s)$  and  $1/\rho(s)$  have the period  $L$  but in general  $F(s)$  has only the period  $\mathcal{C} = NL$ , e.g. if we have one magnet error.

We look for a periodic solution of the inhomogeneous equation (5.5). The general solution is

$$y(s) = aC(s) + bS(s) + u(s) \quad (5.6)$$

$u(s)$  is a special solution of the inhomogeneous equation. It is given by (cf. (3.11))

$$u(s) = S(s) \int_{s_0}^s F(t)C(t)dt - C(s) \int_{s_0}^s F(t)S(t)dt \quad (5.7)$$

The reference point  $s = s_0$  is arbitrary;  $C(s)$  and  $S(s)$  are the cosinelike and sinelike trajectories referred to this point  $s_0$ .

Our goal is to find a periodic solution of the type (5.6)

$$Y(s + \mathcal{C}) = Y(s), \quad Y'(s + \mathcal{C}) = Y'(s) \quad (5.8)$$

Since  $s_0$  is arbitrary we can evaluate the conditions (5.8) at  $s = s_0$  to compute the unknown constants  $a$  and  $b$ .

$$\begin{aligned} ac_1 + bs_1 + u_1 &= ac_0 + bs_0 + u_0 \\ ac'_1 + bs'_1 + u'_1 &= ac'_0 + bs'_0 + u'_0 \end{aligned}$$

Here

$$\begin{aligned} c_0 &= C(s_0) = 1, \quad c'_0 = C'(s_0) = 0 \\ c_1 &= C(s_0 + \mathcal{C}), \quad c'_1 = C'(s_0 + \mathcal{C}) \\ u_0 &= 0, \quad u'_0 = 0, \quad u_1 = u(s_0 + \mathcal{C}) \quad etc. \end{aligned}$$

We obtain

$$a = \frac{s_1 u'_1 - (s'_1 - 1)u_1}{(c_1 - 1)(s'_1 - 1) - c'_1 s_1}$$

The denominator is

$$\begin{aligned} &1 + (c_1 s'_1 - s_1 c'_1) - (c_1 + s'_1) \\ &= 1 + \underbrace{\det \mathbf{M}}_1 - \underbrace{\text{trace } \mathbf{M}}_{2 \cos 2\pi Q} = 4 \sin^2(\pi Q) \end{aligned}$$

Here  $\mathbf{M} = \begin{pmatrix} C(s_0 + \mathcal{C}) & S(s_0 + \mathcal{C}) \\ C'(s_0 + \mathcal{C}) & S'(s_0 + \mathcal{C}) \end{pmatrix}$  is the transformation matrix for one revolution.

The numerator is

$$s_1 \left( s'_1 \int FC - c'_1 \int FS \right) - (s'_1 - 1) \left( s_1 \int FC - c_1 \int FS \right)$$

$$\left( \text{abbreviation: } \int FS = \int_{s_0}^{s_0+Q} F(t)S(t)dt \right)$$

$$\begin{aligned} \text{Num.} &= \underbrace{(c_1 s'_1 - c'_1 s_1)}_{\det M = 1} \int FS - c_1 \int FS + s_1 \int FC \\ c_1 &= C(s_0 + Q) = \cos 2\pi Q + \alpha_0 \sin 2\pi Q \quad (\text{see Eq.(4.42)}) \\ s_1 &= \beta_0 \sin 2\pi Q \end{aligned}$$

$$\begin{aligned} \text{Num.} &= (1 - \cos 2\pi Q - \alpha_0 \sin 2\pi Q) \sqrt{\beta_0} \int_{s_0}^{s_0+Q} \sqrt{\beta(t)} F(t) \sin \Delta\Phi(t) dt \\ &\quad + \sqrt{\beta_0} \sin 2\pi Q \int_{s_0}^{s_0+Q} \sqrt{\beta(t)} F(t) (\cos \Delta\Phi(t) + \alpha_0 \sin \Delta\Phi(t)) dt \\ &= \sqrt{\beta_0} 2 \sin \pi Q \int_{s_0}^{s_0+Q} \sqrt{\beta(t)} F(t) \cos(\Phi(t) - \Phi_0 - \pi Q) dt \end{aligned}$$

Now  $a = Y(s_0)$ . The point  $s_0$  was chosen arbitrarily. Therefore we get for the closed trajectory

$$\boxed{\begin{aligned} Y(s) &= \frac{\sqrt{\beta(s)}}{2 \sin \pi Q} \int_s^{s+Q} \sqrt{\beta(t)} F(t) \cos(\Phi(t) - \Phi(s) - \pi Q) dt \\ &= \frac{\sqrt{\beta(s)}}{2 \sin \pi Q} \oint \sqrt{\beta(t)} F(t) \cos(|\Phi(t) - \Phi(s)| - \pi Q) dt \end{aligned}} \quad (5.9)$$

Note that in the first equation  $t > s$  is required, i.e. the integration has to start at  $s$ , while in the second one this restriction does not apply.

The closed periodic dispersion function is obtained for  $F(t) = \frac{1}{\rho(t)} \cdot \frac{\Delta p}{p_0}$ :

$$\boxed{D(s) \equiv \eta(s) = \frac{\sqrt{\beta(s)}}{2 \sin \pi Q} \oint \frac{\sqrt{\beta(t)}}{\rho(t)} \cos(|\Phi(t) - \Phi(s)| - \pi Q) dt} \quad (5.10)$$

These equations exhibit an essential instability of a circular accelerator: a finite dispersion exists only if the number of betatron oscillations per revolution  $Q$  is different from an integer. Consider an integer  $Q$  and look at a certain dipole. A particle with  $\Delta p$  will receive a different kick angle than the reference particle with  $p = p_0$ . Since  $Q$  is integer, these angular deviations add up coherently from turn to turn, and soon the particle hits the vacuum chamber.

While Eqs. (5.9) and (5.10) are very useful for understanding the overall oscillatory behaviour of closed orbits in the presence of dipole errors or momentum errors, numerical calculations will be easier if matrix formalism is applied. This is provided by Eq. (3.10), since  $x_D(s) = \eta(s) \cdot \frac{\Delta p}{p_0}$  is a possible particle trajectory (we use now  $D$ ,  $D'$  for the dispersion trajectory part of the transfer matrix and  $\eta$ ,  $\eta'$  for the periodic dispersion).

$$\begin{pmatrix} \eta \frac{\Delta p}{p_0} \\ \eta' \frac{\Delta p}{p_0} \\ \frac{\Delta p}{p_0} \end{pmatrix} = \begin{pmatrix} C & S & D \\ C' & S' & D' \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \eta \frac{\Delta p}{p_0} \\ \eta' \frac{\Delta p}{p_0} \\ \frac{\Delta p}{p_0} \end{pmatrix}$$

or

$$\begin{pmatrix} \eta \\ \eta' \\ 1 \end{pmatrix} = \begin{pmatrix} C & S & D \\ C' & S' & D' \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \eta \\ \eta' \\ 1 \end{pmatrix}$$

This means:  $\eta$  and  $\eta'$  are eigenvector components of the revolution matrix  $M$  for the eigenvalue 1. (We leave the proof that the  $3 \times 3$  matrix  $M$  always has an eigenvalue 1 as an exercise.)

Explicitly, the last equation implies

$$\eta = \frac{(1 - S')D + SD'}{2 - C - S'} = \frac{(1 - S')D + SD'}{4 \sin^2 \pi Q}$$

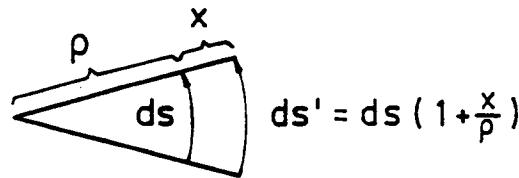
$$\eta' = \frac{C'D + (1 - C)D'}{4 \sin^2 \pi Q}$$

Again it is seen that integer  $Q$  values are to be avoided. Once  $\eta$  and  $\eta'$  have been determined at one point  $s_0$ , the values at any other point  $s$  are calculated by a simple matrix multiplication:

$$\begin{pmatrix} \eta \\ \eta' \\ 1 \end{pmatrix}_s = M(s/s_0) \begin{pmatrix} \eta \\ \eta' \\ 1 \end{pmatrix}_{s_0}$$

#### Momentum compaction

A particle with  $\Delta p/p_0 > 0$  travels on a revolution a longer distance than the reference particle ( $\Delta p = 0$ ).



Consider a particle moving on the closed dispersion trajectory.

$$x_D(s) = D(s) \frac{\Delta p}{p_0}$$

The circumference for this particle is

$$C' = \oint \left( 1 + \frac{x_D(s)}{\rho} ds \right) = C + \Delta C$$

$$\frac{\Delta C}{C} = \alpha \frac{\Delta p}{p_0} \quad (5.11)$$

$\alpha = \frac{1}{C} \oint \frac{D(s)}{\rho(s)} ds$

(5.12)

The quantity  $\alpha$  is the relative change in orbit length divided by the relative momentum deviation. It is called “momentum compaction factor”, which is a rather misleading notion. A rough estimate in terms of the horizontal  $Q$ -value is given by

$$\alpha \approx \frac{1}{Q_x^2}$$

## 5.2 Dispersion in transfer lines

In transfer lines, the dispersion function is also derived from Eqs. (5.6) and (5.7), but periodicity cannot be assumed anymore. Instead, initial conditions are derived from the beam properties at the entrance of the transfer lines. If here a significant correlation exists between the phase space coordinates of the particles and their respective momenta, appropriate initial conditions are

$$D_0 \left\langle \left( \frac{\delta p_i}{p} \right)^2 \right\rangle_i = \left\langle x_i \frac{\delta p_i}{p} \right\rangle_i; \quad D'_0 \left\langle \left( \frac{\delta p_i}{p} \right)^2 \right\rangle_i = \left\langle x'_i \frac{\delta p_i}{p} \right\rangle_i;$$

where  $\langle \rangle_i$  denotes averaging over the whole ensemble of particles at the entrance.

The dispersion as a function of the position  $s$  is then

$$D(s) = D_0 C(s) + D'_0 S(s) + S(s) \int_{s_0}^s \frac{1}{\rho(t)} C(t) dt - C(s) \int_{s_0}^s \frac{1}{\rho(t)} S(t) dt \quad (5.13)$$

If no such correlation is known at the entrance, one usually chooses  $D_0 = D'_0 = 0$ , see Eq. (3.11). A transfer line is called *non-dispersive* if

$$D_0 = D'_0 = 0, \quad \int_{s_0}^s \frac{C(t)}{\rho(t)} dt = 0 \quad \int_{s_0}^s \frac{S(t)}{\rho(t)} dt = 0$$

It is instructive to consider also the generalization of the momentum compaction factor. The relative change in orbit length per relative momentum deviation is given by

$$\alpha(s, s_0) = \frac{\Delta L/L_0}{\Delta p/p_0} = \frac{1}{L_0} \int_{s_0}^s \frac{D(t)}{\rho(t)} dt \quad \text{with} \quad L_0 = \int_{s_0}^s dt$$

Finally we consider the relative change in time of flight per relative momentum deviation:

$$\frac{\Delta t/t_0}{\Delta p/p_0} = \frac{p_0}{t_0} \frac{\Delta(L/v)}{\Delta p} = \alpha(s, s_0) - 1 + \frac{v^2}{c^2}$$

This quantity is sometimes denoted by  $\eta(s, s_0)$ . If  $\eta(s, s_0) = 0$ , the transfer line is called *isochronous*, because then the time of flight does not depend on momentum. For ultrarelativistic particles ( $(1 - v^2/c^2) \rightarrow 0$ ), there is no difference between  $\alpha$  and  $\eta$ .

## 5.3 Influence of field errors

In this section we want to touch very briefly the effect of a dipole or quadrupole error.

### a) Dipole error

If  $B_z = B_0 + \Delta B$  or  $B_x = \Delta B$ , we have an additional,  $s$ -dependent Lorentz force. This leads again to an equation of the type (5.5). The periodic closed orbit is given by (5.9) with  $F(t) = \frac{e}{p_0} \Delta B(t)$ . In order to have bounded motion the  $Q$  value must be non integer,  $Q \neq n$ .

We see that even for particles with reference momentum  $p_0$  an integer  $Q$  value is forbidden, since small field errors are always present.

The main source of orbit errors in accelerators is the displacement of quadrupoles from the design orbit. Correction dipoles are needed to correct the orbit. Eq. (5.9) tells us that their effect is largest if they are placed at points with a large beta function. So a horizontally deflecting correction dipole should be placed close to a horizontally focusing quadrupole and a vertically deflecting correction dipole close to a vertically focusing quadrupole.

b) Quadrupole errors

Let  $K_0(s)$  be the design quadrupole strength,  $\Delta K(s)$  the error. For a particle with  $\Delta p = 0$  we have

$$y'' + (K_0(s) + \Delta K(s))y = 0$$

Call  $M_0$  the transformation matrix for a revolution in the undisturbed machine

$$M_0 = I \cos \mu_0 + J \sin \mu_0$$

$$\mu_0 = 2\pi Q$$

Suppose the error occurs only at  $s = s_1$  over a short length  $ds_1$  and let  $M$  denote the transformation matrix for the disturbed machine.

We get

$$M = mm_0^{-1}M_0$$

Here  $m_0$  is the matrix of the section of length  $ds_1$  in the undisturbed case,  $m$  the matrix in the disturbed case

$$m_0 = \begin{pmatrix} 1 & 0 \\ -K_0(s_1)ds_1 & 1 \end{pmatrix} \quad m = \begin{pmatrix} 1 & 0 \\ (-K_0(s_1) + \Delta K(s_1))ds_1 & 1 \end{pmatrix}$$

To first order in  $ds_1$ :

$$M = \begin{pmatrix} 1 & 0 \\ -\Delta K ds_1 & 1 \end{pmatrix} \cos \mu_0 + \begin{pmatrix} \alpha & \beta \\ -\alpha \Delta K ds_1 - \gamma & -\beta \Delta K ds_1 - \alpha \end{pmatrix} \sin \mu_0$$

$$\cos \mu = \frac{1}{2} \text{trace } M = \cos \mu_0 - \frac{1}{2} \beta \Delta K ds_1 \sin \mu_0$$

If the perturbation is distributed around the ring, we get

$$\Delta(\cos \mu) = \cos \mu - \cos \mu_0 = -\frac{\sin \mu_0}{2} \oint \beta(s) \Delta K(s) ds$$

Now

$$\Delta Q = \frac{\Delta \mu}{2\pi} = -\frac{\Delta(\cos \mu)}{2\pi \sin \mu_0}$$

A gradient error therefore leads to a shift in the  $Q$  value

$$\Delta Q = \frac{1}{4\pi} \oint \beta(s) \Delta K(s) ds \quad (5.14)$$

The  $Q$  shift is proportional to both the magnitude of the gradient error and the beta function at the location of the error.

A gradient error changes the beta function itself. Without proof we state the result

$$\Delta \beta(s) = \frac{\beta(s)}{2 \sin(2\pi Q)} \oint \beta(t) \Delta K(t) \cos(2|\Phi(t) - \Phi(s)| - 2\pi Q) dt \quad (5.15)$$

Again the change of the beta function is proportional to the magnitude of the perturbation and the amplitude of the beta function itself at the point of the perturbation.

Gradient errors in the interaction region quadrupoles are therefore most dangerous. The second important result is that  $\Delta \beta$  remains only finite if  $Q$  is different from a half-integer number ( $\sin(2\pi Q)$  in the denominator).

The observation that dipole errors lead to integer resonances, quadrupole errors to resonances at half-integer  $Q$  values indicates that sextupole fields excite resonances at third-integer  $Q$  values. This is in fact the case but not the subject of the present introductory course.<sup>2</sup> Moreover, one has in general a coupling between horizontal and vertical betatron oscillations due to sextupoles or misalignment of magnets (skew quadrupole fields, see the table at the end of section 2f) etc. This leads to a more general condition

$$mQ_x + nQ_z \neq \ell \quad (5.16)$$

( $m, n, \ell$  integer numbers;  $m, n$  small)

The working point  $(Q_x, Q_z)$  has to be chosen in a reasonable distance from the resonance lines (Fig. 34)

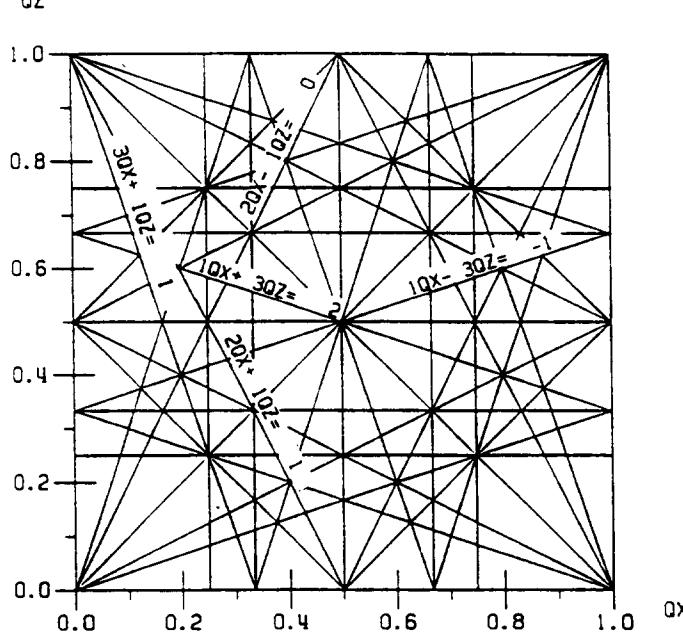


Figure 34: Resonance diagram up to fourth order. Some of the resonance lines have been identified explicitly.

<sup>2</sup>Resonances are discussed in detail in E. Wilson's lectures at this and the previous accelerator schools (see e.g. the proceedings of the 1984 accelerator school, CERN 85-19).

## 5.4 Chromaticity

Particles with  $\Delta p \neq 0$  are focused differently in the quadrupoles. This leads to a shift of the  $Q$  value.

From

$$K = \frac{eg}{p}$$

we get

$$\Delta K = \frac{dK}{dp} \Delta p = -\frac{eg}{p_0} \cdot \frac{\Delta p}{p_0} = -K_0 \frac{\Delta p}{p_0}$$

Formally this is the same as a gradient error. We can therefore use (5.14) to calculate the  $Q$  shift

$$\Delta Q = -\frac{1}{4\pi} \oint \beta(s) K(s) ds \frac{\Delta p}{p_0} = \xi \frac{\Delta p}{p_0}$$

$$\boxed{\xi = -\frac{1}{4\pi} \oint \beta(s) K(s) ds} \quad (5.17)$$

$\xi$  is called the chromaticity of the machine. For a linear magnet lattice it is always negative. The main contribution to the chromaticity comes from quadrupoles which are strongly excited and where the  $\beta$  function is large (e.g. interaction region quadrupoles).

In big accelerators the chromaticity arising from the linear lattice (also called the “linear” or “natural” chromaticity) is a large quantity (e.g.  $\xi \approx -60$  in the HERA storage rings). Then, the “tune” spread due to the finite momentum band becomes so large that some part of the beam unavoidably hits dangerous resonance lines. In the HERA case, for instance, a beam with a relative momentum spread of only  $\pm 10^{-3}$  would cover a tune range of 0.12 ! For this reason, and in order to avoid the so-called “head-tail” instability, one has to compensate the chromaticity. This can be achieved with sextupoles. The sextupole magnets have to be placed at locations where the closed dispersion orbit  $D(s)$  is nonzero, (Fig. 35).

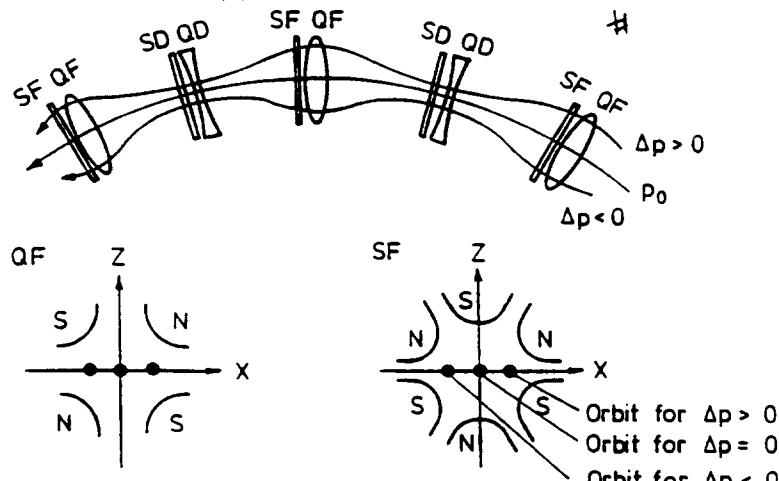


Figure 35: Dispersion trajectory in a sextupole magnet

Consider a particle with  $\Delta p \neq 0$  moving without betatron oscillation on the closed dispersion trajectory.

$$x_D(s) = D(s) \frac{\Delta p}{p_0}$$

Now put a sextupole magnet at a place with  $D(s) \neq 0$ .

The sextupole field at  $x = x_D$  in the horizontal plane  $z = 0$  is

$$B_z = \frac{1}{2} g' x^2 \quad (5.18)$$

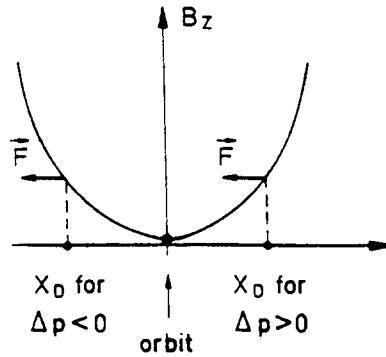


Figure 36: Force on electrons in a sextupole magnet

Particles which move with  $\Delta p = 0$  on the reference orbit are not influenced. With the polarity shown in Fig. 35, the sextupole deflects electrons with  $\Delta p > 0$  towards the central orbit and electrons with  $\Delta p < 0$  away from it (Fig. 36). It is therefore clear that the closed dispersion orbit (5.10), which was computed for a “linear” machine with only dipole and quadrupole magnets, will change when sextupoles are switched on. The dispersion orbit of the “nonlinear” machine can be determined by an iterative procedure.

Now consider a particle that conducts betatron oscillations around that orbit. We want to show that the sextupole acts like a position-dependent quadrupole and influences both horizontal and vertical motion.

Let  $\Delta p > 0$  and consider small deviations  $\tilde{x}, \tilde{z}$  from the closed dispersion trajectory

$$\begin{aligned} x &= x_D + \tilde{x} \\ z &= 0 + \tilde{z} \end{aligned}$$

To first order in the small quantities  $\tilde{x}, \tilde{z}$ :

$$\begin{aligned} B_x &\approx g' x_D \tilde{z} = \left( g' D(s) \frac{\Delta p}{p_0} \right) \tilde{z} \\ B_z &\approx \underbrace{\frac{1}{2} g' x_D^2}_{\text{causes deflection of}} \quad + \left( g' D(s) \frac{\Delta p}{p_0} \right) \tilde{x} \end{aligned} \quad (5.19)$$

closed dispersion orbit

Thus the sextupole acts like a quadrupole whose strength increases linearly with  $\Delta p$ . The equivalent quadrupole strength is

$$\frac{eg'x_D}{p_0} = m \cdot D(s) \frac{\Delta p}{p_0} \quad (5.20)$$

$$m = \frac{eg'}{p_0} \quad (\text{Eq. (2.13)})$$

In a regular FODO lattice, two families of sextupoles are needed, (horizontally) focusing ones near the F quadrupoles and defocusing ones near the D quadrupoles. By a proper choice of their strengths one can achieve simultaneous compensation of the horizontal and vertical chromaticities.

$$\xi_{comp} = -\frac{1}{4\pi} \oint [K(s) - m(s)D(s)] \beta(s) ds \quad (5.21)$$

The sextupole compensation of the chromaticity is unavoidable in large rings but has unfortunate consequences: one introduces non-linear fields and a coupling between horizontal and vertical motion (see (2.12)). These adverse effects can be minimized by using a large number of sextupoles with moderate strength, distributed around the ring, rather than a few very strong sextupoles.

## 6 COMPARISON OF STRONG AND WEAK FOCUSING

### 6.1 Simplified model of strong focusing

In this chapter we follow closely K. Steffen's lectures at the Orsay accelerator school. We consider a FODO channel and make the following simplifying assumptions

- The weak focusing of the bending magnets is neglected since the radius of curvature is much larger than the focal length of the quadrupoles. (HERA p-ring:  $\rho = 584$  m,  $f = 16.4$  m)
- we neglect the short drift spaces between dipoles and quadrupoles
- $F$  and  $D$  quadrupoles have the same strength  $1/f$  and are treated as thin lenses ( $\ell_q \rightarrow 0$ ).

The simplified model depends only on three parameters

$$\ell = L/2 \text{ half cell length} \simeq \text{dipole length}$$

$$\frac{1}{\rho} \text{ bending strength; } \varphi = \frac{\ell}{\rho}.$$

$$\pm \frac{1}{f'} = \pm \frac{1}{2f} \text{ strength of } \underline{\text{half}} \text{ quadrupole}$$

Its periodic solution and individual particle trajectories have been illustrated in Figs. 28, 30, 31. We now want to explicitly calculate the Twiss parameters as a function of  $\ell, \rho, f$ . We compute the transformation through a half cell from the middle of the  $F$  quadrupole to the middle of the  $D$  quadrupole (Fig. 37)

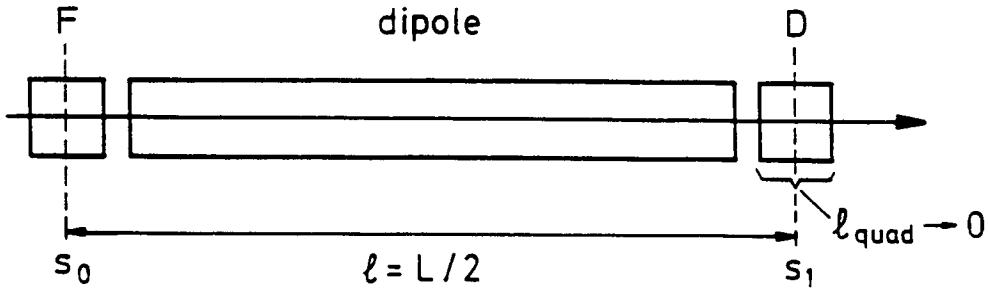


Figure 37: Half cell

$$\begin{aligned}
 M(s_1/s_0) &= \begin{pmatrix} 1 & 0 \\ \frac{1}{f'} & 1 \end{pmatrix} \begin{pmatrix} 1 & \rho \sin \varphi \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -\frac{1}{f'} & 1 \end{pmatrix} \\
 &\text{---} \quad \frac{1}{2} \text{ D-quad.} \quad \text{rectangular dipole} \quad \frac{1}{2} \text{ F-quad.} \\
 M(s_1/s_0) &= \begin{pmatrix} 1 - \frac{\rho}{f'} \sin \varphi & \rho \sin \varphi \\ -\frac{\rho}{f'^2} \sin \varphi & 1 + \frac{\rho}{f'} \sin \varphi \end{pmatrix} \tag{6.1}
 \end{aligned}$$

The transformation through the next half cell to the middle of the next *F* quadrupole at  $s_2 = s_1 + \ell$  is simply obtained by replacing  $f'$  by  $-f'$  in (6.1).

So the transfer matrix for the whole cell is

$$\begin{aligned}
 \mathbf{M}(s_0) &= M(s_0 + L/s_0) = M(s_2/s_1) \cdot M(s_1/s_0) \\
 \mathbf{M}(s_0) &= \begin{pmatrix} 1 - 2\frac{\rho^2}{f'^2} \sin^2 \varphi & 2\rho \sin \varphi + 2\frac{\rho^2}{f'} \sin^2 \varphi \\ -2\frac{\rho}{f'^2} \sin \varphi + 2\frac{\rho^2}{f'^3} \sin^2 \varphi & 1 - 2\frac{\rho^2}{f'^2} \sin^2 \varphi \end{pmatrix} \\
 \cos \mu &= \frac{1}{2} \text{trace } \mathbf{M} = 1 - 2\frac{\rho^2}{f'^2} \sin^2 \varphi \tag{6.2}
 \end{aligned}$$

Now

$$\begin{aligned}
 \varphi &= \frac{\ell}{\rho} \ll 1, \text{ so } \sin \varphi \approx \frac{\ell}{\rho}; \quad \frac{1}{f'} = \frac{1}{2f} \\
 \cos \mu &= 1 - \frac{1}{2} \frac{\ell^2}{f'^2} = 1 - \frac{1}{8} \frac{L^2}{f^2}
 \end{aligned}$$

So in the approximation  $\sin \varphi \approx \varphi$  we get the same result for the phase advance in a FODO cell as in section 4.7 where we treated the bending magnets as drift spaces. The present treatment offers additional information. In the middle of an *F* quadrupole the beta function assumes its maximum value  $\hat{\beta}$ , in the middle of a *D* quadrupole its minimum value  $\check{\beta}$ . The slope  $\beta'$  is zero here, so  $\alpha = 0$  and  $\gamma = 1/\beta$ . The matrix (4.42), applied to the transformation  $s_0 \rightarrow s_1$ , is therefore (with  $\Delta\Phi = \mu/2$ )

$$\begin{pmatrix} C & S \\ C' & S' \end{pmatrix} = \begin{pmatrix} \sqrt{\check{\beta}/\hat{\beta}} \cos \Delta\Phi & \sqrt{\hat{\beta}\check{\beta}} \sin \Delta\Phi \\ -\frac{1}{\sqrt{\hat{\beta}\check{\beta}}} \sin \Delta\Phi & \sqrt{\hat{\beta}/\check{\beta}} \cos \Delta\Phi \end{pmatrix} \quad \Delta\Phi = \frac{\mu}{2} \quad (6.3)$$

From the equality of the matrices (6.1) and (6.3) it follows

$$\frac{S'}{C} = \hat{\beta}/\check{\beta} = \frac{1 + \frac{\rho}{f'} \sin \varphi}{1 - \frac{\rho}{f'} \sin \varphi}$$

$$-\frac{S}{C'} = \hat{\beta}\check{\beta} = f'^2$$

Now  $f' = 2f$  and from (6.2)  $\frac{\rho}{f'} \sin \varphi = \sin \frac{\mu}{2} \approx \frac{L}{4f} = \frac{\varepsilon}{f'}$ . This allows to compute  $\hat{\beta}$  and  $\check{\beta}$  as functions of  $\mu$

$$\hat{\beta} = L \frac{1 + \sin \frac{\mu}{2}}{\sin \mu}$$

$$\check{\beta} = L \frac{1 - \sin \frac{\mu}{2}}{\sin \mu} \quad (6.4)$$

We see that our basic assumption  $|\cos \mu| < 1$ , that is  $\sin \mu \neq 0$ , is necessary to get a finite value for  $\hat{\beta}$ .

For  $0 < \mu < \pi$ :  $\sin \frac{\mu}{2} < 1$ ,  $\frac{L}{4f} < 1$  (see (4.45)). For a given cell length  $L$ , the phase advance with the smallest  $\hat{\beta}$  requires the least beam aperture. We obtain it by differentiating (6.4):

$$\text{From } \frac{d\hat{\beta}}{d\mu} = 0 \quad \text{it follows} \quad \sin^2 \frac{\mu}{2} \left( 2 + \sin \frac{\mu}{2} \right) = 1$$

$$\mu = 76.35^\circ$$

In practice, phase advances per cell are typically  $60^\circ$  to  $90^\circ$ .

To transform the dispersion through the half cell we have to compute the  $3 \times 3$  transfer matrix.

$$\begin{pmatrix} C & S & D \\ C' & S' & D' \\ 0 & 0 & 1 \end{pmatrix}_{\text{half cell}} = \begin{pmatrix} 1 & 0 & 0 \\ \frac{1}{f'} & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}_{\frac{1}{2} \text{ D quad.}} \begin{pmatrix} 1 & \rho \sin \varphi & \rho(1 - \cos \varphi) \\ 0 & 1 & 2 \tan \frac{\varphi}{2} \\ 0 & 0 & 1 \end{pmatrix}_{\text{dipole}} \begin{pmatrix} 1 & 0 & 0 \\ -\frac{1}{f'} & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}_{\frac{1}{2} \text{ F quad.}}$$

$$\begin{pmatrix} C & S & D \\ C' & S' & D' \\ 0 & 0 & 1 \end{pmatrix}_{\text{half cell}} = \begin{pmatrix} 1 - \frac{\rho}{f'} \sin \varphi & \rho \sin \varphi & \rho(1 - \cos \varphi) \\ -\frac{\rho}{f'^2} \sin \varphi & 1 + \frac{\rho}{f'} \sin \varphi & \frac{\rho}{f'}(1 - \cos \varphi) + 2 \tan \frac{\varphi}{2} \\ 0 & 0 & 1 \end{pmatrix}$$

Now in the center of the  $F$ -quad:  $D(s_0) = \hat{D}$ ,  $D'(s_0) = 0$  and in the center of the  $D$ -quad:  $D(s_1) = \check{D}$ ,  $D'(s_1) = 0$

So

$$\begin{pmatrix} C_1 & S_1 & \check{D} \\ C'_1 & S'_1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} C & S & D \\ C' & S' & D' \\ 0 & 0 & 1 \end{pmatrix}_{\text{half cell}} \begin{pmatrix} C_0 & S_0 & \hat{D} \\ C'_0 & S'_0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

This yields two equations for  $\hat{D}$  and  $\check{D}$ . The result is

$$\begin{aligned} \hat{D} &= \frac{\ell^2}{\rho} \frac{1 + \frac{1}{2} \sin \frac{\mu}{2}}{\sin^2 \frac{\mu}{2}} \\ \check{D} &= \frac{\ell^2}{\rho} \frac{1 - \frac{1}{2} \sin \frac{\mu}{2}}{\sin^2 \frac{\mu}{2}} \end{aligned} \quad (6.5)$$

The highest and lowest values of the beta function and the dispersion are plotted in Fig. 38 as a function of the phase advance  $\mu$ .

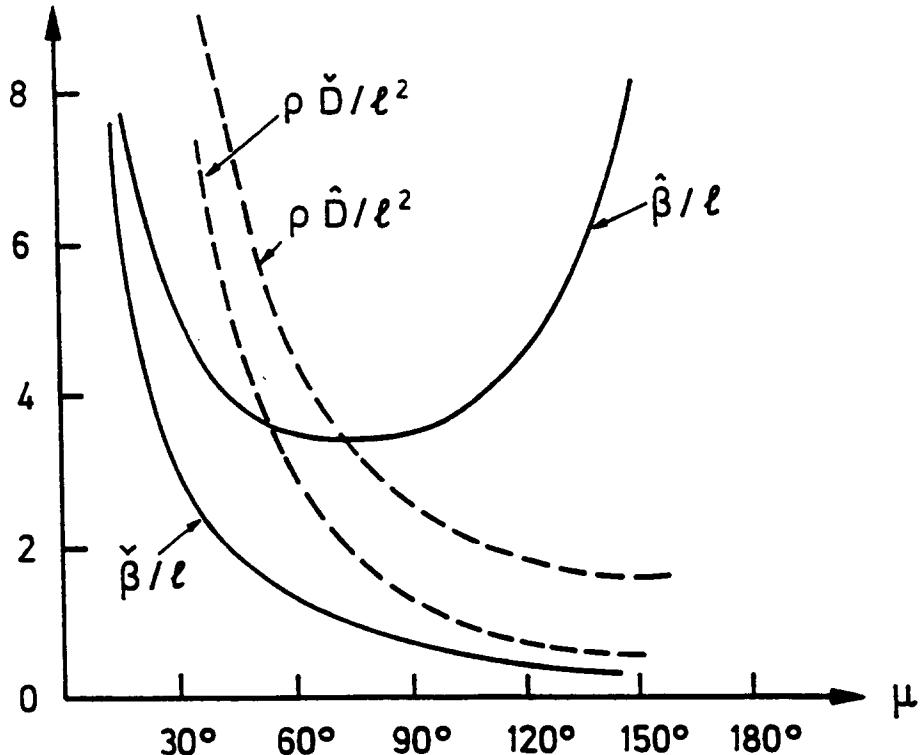


Figure 38: The extreme values of  $\beta$  and  $D$  as functions of the phase advance  $\mu$  in the cell.

It is instructive to compute  $\beta$  and  $D$  as functions of  $s$  within the half cell. For simplicity we set  $s_0 = 0$ . Starting from the center of the  $F$ -quad at  $s_0 = 0$  we first have to apply the matrix equation (4.40) for a half quadrupole to compute  $\alpha, \beta, \gamma$  at the exit of the  $F$ -quad and then the same equation for a bending magnet of length  $s$  with  $0 < s < \ell$ .

$$\begin{pmatrix} \beta \\ \alpha \\ \gamma \end{pmatrix}_{\text{exit F-quad}} = \begin{pmatrix} 1 & 0 & 0 \\ \frac{1}{f'} & 1 & 0 \\ \frac{1}{f'^2} & \frac{2}{f'} & 1 \end{pmatrix} \begin{pmatrix} \hat{\beta} \\ 0 \\ 1/\hat{\beta} \end{pmatrix} = \begin{pmatrix} \hat{\beta} \\ \frac{\hat{\beta}}{f'} \\ \frac{\hat{\beta}}{f'^2} + \frac{1}{\hat{\beta}} \end{pmatrix}$$

matrix (4.40)      center of  
for 1/2 quad      F-quad

Now for a bending magnet of length  $s$

$$\begin{pmatrix} C & S \\ C' & S' \end{pmatrix} = \begin{pmatrix} 1 & \rho \sin \varphi \\ 0 & 1 \end{pmatrix} \approx \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix} \quad (\text{for } \varphi = \frac{s}{\rho} \ll 1)$$

so

$$\begin{pmatrix} \beta(s) \\ \alpha(s) \\ \gamma(s) \end{pmatrix} = \begin{pmatrix} 1 & -2s & s^2 \\ 0 & 1 & -s \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \hat{\beta} \\ \frac{\hat{\beta}}{f'} \\ \frac{\hat{\beta}}{f'^2} + \frac{1}{\hat{\beta}} \end{pmatrix}$$

matrix (4.40)  
for dipole

We get

$$\beta(s) = \hat{\beta} - 2\frac{\hat{\beta}}{f'}s + \left(\frac{\hat{\beta}}{f'^2} + \frac{1}{\hat{\beta}}\right)s^2 \quad (6.6)$$

$$\rightarrow \beta(s) = \hat{\beta} \text{ for } s = 0, \quad \beta(s) = \check{\beta} \text{ for } s = \ell.$$

The dispersion  $D(s)$  is obtained using the transformation matrices (3.15) and (3.20).

$$\begin{pmatrix} C & S & D \\ C' & S' & D' \\ 0 & 0 & 1 \end{pmatrix}_{\text{exit F-quad}} = \begin{pmatrix} 1 & 0 & 0 \\ -\frac{1}{f'} & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} C_{\circ} & S_{\circ} & \hat{D} \\ C'_{\circ} & S'_{\circ} & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

1/2 F-quad      values in center  
of F-quad

$$\begin{pmatrix} C(s) & S(s) & D(s) \\ C'(s) & S'(s) & D'(s) \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & s & \frac{1}{2}\frac{s^2}{\rho} \\ 0 & 1 & s/\rho \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} C & S & D \\ C' & S' & D' \\ 0 & 0 & 1 \end{pmatrix}_{\text{exit F-quad}}$$

matrix (3.20)  
for  $\varphi = \frac{s}{\rho} \ll 1$

The result is

$$D(s) = \hat{D} \left( 1 - \frac{s}{f'} \right) + \frac{1}{2} \frac{s^2}{\rho} \quad (6.7)$$

$$\Rightarrow D(s_0 = 0) = \hat{D}, \quad D(s = \ell) = \check{D}$$

In an accelerator, where the arcs consist of identical FODO cells, the momentum compaction factor is

$$\alpha = \frac{1}{2\pi R} \oint \frac{D(s)}{\rho(s)} ds \approx \frac{1}{\rho} \cdot \frac{2\pi\rho}{2\pi R} \cdot \frac{1}{2} (\hat{D} + \check{D}) = \frac{\hat{D} + \check{D}}{2R} = \frac{\ell^2}{\rho R \sin^2 \frac{\mu}{2}} = \frac{f'^2}{\rho R}$$

$R = \frac{C}{2\pi}$  is the average machine radius including the straight sections.

For example

$$\alpha = \frac{f'^2}{R\rho} = \frac{4f^2}{R\rho} \approx 1.4 \cdot 10^{-3} \quad \text{for the HERA proton ring} \quad (6.8)$$

The chromaticity can be estimated as follows (cf. Eq. (5.17)):

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

In a focusing quadrupole:

$$\beta \approx \hat{\beta}, \quad K(s)\Delta s = \frac{1}{f} \quad (\Delta s \text{ denotes the length of the quad});$$

in a defocusing quadrupole:

$$\beta \approx \check{\beta}, \quad K\Delta s = -\frac{1}{f}$$

Therefore, the chromaticity caused by the regular cell structure of the accelerator is

$$\xi \approx -\frac{1}{4\pi} N \frac{1}{f} (\hat{\beta} - \check{\beta}) = -\frac{N}{\pi} \tan \frac{\mu}{2} \quad N = \text{number of cells} \quad (6.9)$$

Each FODO cell contributes  $-\frac{1}{\pi} \tan \frac{\mu}{2}$ .

In the HERA proton ring, the contribution of the arcs is  $N = 104$ ,  $\hat{\beta} \approx 80m$ ,  $\check{\beta} \approx 20m$ ,  $f \approx 16m$ ; so  $\xi_{\text{arcs}} \approx -30$ .

The interaction region quadrupoles contribute a large amount to the total chromaticity since there  $\beta(s)$  and  $K(s)$  are both large. The chromaticity of the straight sections in the HERA proton ring (including the interaction regions) amounts to  $\xi_{\text{straight sections}} \approx -30$ . Thus the total linear chromaticity of the HERA proton ring is  $\xi_{\text{HERA}} \approx -60$ .

For a particle with a momentum error of  $\Delta p/p_0 = 5 \cdot 10^{-4}$  this would imply a tune shift of  $\Delta Q = -0.03$  if no sextupoles were present.

## 6.2 Stability diagram

In principle one can choose different strengths for the focusing and the defocusing quadrupoles. Call

$$\frac{\ell}{f'_1} = F > 0, \text{ and } -\frac{\ell}{f'_2} = D > 0$$

Then for the horizontal motion, we have(cf. Eq. (3.21))

$$\begin{pmatrix} C & S \\ C' & S' \end{pmatrix}_{\text{half cell}} = \begin{pmatrix} 1-F & \ell \\ -\frac{1}{t}(F-D+FD) & 1+D \end{pmatrix} \quad (6.10)$$

This matrix has to be equal to the matrix (6.3). In particular

$$-C'S = F - D + FD = \sin^2 \frac{\mu_x}{2}$$

For the vertical motion, one simply has to interchange  $F$  and  $D$  in (6.10)

$$D - F + FD = \sin^2 \frac{\mu_z}{2}$$

Now  $0 \leq \sin^2 \frac{\mu}{2} \leq 1$  has to hold in both cases. The criterion for stability in both planes is therefore

$$\begin{aligned} 0 \leq F - D + FD \leq 1 \\ 0 \leq D - F + FD \leq 1 \end{aligned}$$

The limits of the stable region are

$$\begin{aligned} \sin \frac{\mu_x}{2} = 0 \Rightarrow F = \frac{D}{1+D} & \quad \sin \frac{\mu_z}{2} = 0 \Rightarrow D = \frac{F}{1+F} \\ \sin \frac{\mu_x}{2} = 1 \Rightarrow F = 1 & \quad \sin \frac{\mu_z}{2} = 1 \Rightarrow D = 1 \end{aligned}$$

The stable region has the appearance of a necktie (Fig. 39).

## 6.3 Weak focusing with constant gradient

Let us assume a continuous ring magnet with the magnetic field independent of  $s$ . The focusing strengths

$$\begin{aligned} K_z &= k \\ K_x &= \left( \frac{1}{\rho^2} - k \right) \end{aligned} \quad (6.11)$$

can be made both positive, if  $k$  is smaller than the “weak focusing term”  $\frac{1}{\rho^2}$ .

We choose a small gradient  $g$

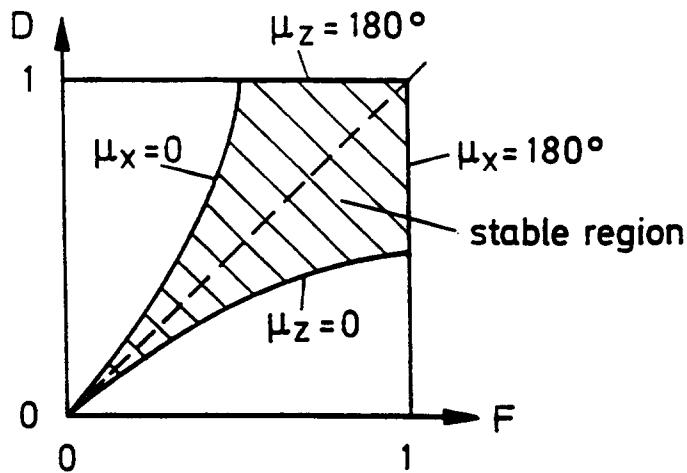


Figure 39: Stability diagram for different strengths of focusing and defocusing quadrupoles

$$B_z(x) = B_0 - gx \quad (6.12)$$

Then the field index  $n$ , defined in Eq. (1.5), is

$$n = -\frac{\rho}{B_0} \frac{\partial B_z}{\partial x} = \frac{\rho g}{B_0} = \frac{p_0 \rho}{e B_0} k = \rho^2 k \quad (6.13)$$

$$K_x = \frac{1-n}{\rho^2}, \quad K_z = \frac{n}{\rho^2} \quad (6.14)$$

From the conditions  $K_x > 0$ ,  $K_z > 0$  for focusing in both planes we recover our old result Eq. (1.6):  $0 < n < 1$ .

The transfer matrix for one revolution is for the horizontal motion

$$\begin{pmatrix} C & S & D \\ C' & S' & D' \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} \cos \varphi & \frac{\rho}{\sqrt{1-n}} \sin \varphi & \frac{\rho}{1-n} (1 - \cos \varphi) \\ -\frac{\sqrt{1-n}}{\rho} \sin \varphi & \cos \varphi & \frac{1}{\sqrt{1-n}} \sin \varphi \\ 0 & 0 & 1 \end{pmatrix} \quad (6.15)$$

The matrix (6.15) does not depend on the start point  $s$ . If we equate it to the  $(3 \times 3)$ -extension of the matrix (4.42), we see that  $\beta$  does not depend on  $s$  and therefore  $\alpha = -\frac{1}{2}\beta' = 0$ .

We get

$$\begin{aligned} \beta_x &= \frac{\rho}{\sqrt{1-n}} = \text{const} \\ Q_x &= \sqrt{1-n} \end{aligned} \quad (6.16)$$

The number of betatron oscillations per turn is less than 1.

The dispersion is

$$D = \frac{\rho}{1-n} = \text{const} \quad (6.17)$$

The closed dispersion curve is simply a circle which is concentric with the reference orbit  $r = \rho$ . The momentum compaction factor is therefore

$$\alpha = \frac{1}{2\pi\rho} \oint \frac{D}{\rho} ds = \frac{1}{1-n}$$

For the vertical plane, we replace  $1 - n$  by  $n$

$$\begin{aligned}\beta_z &= \frac{\rho}{\sqrt{n}} = const \\ Q_z &= \sqrt{n}\end{aligned}\tag{6.18}$$

For  $n = \frac{1}{2}$  we have  $\beta_x = \beta_z = \rho\sqrt{2}$  and  $D = 2\rho$ . The following table compares weakly and strongly focusing machines for  $\rho = 25 m$  and  $\rho = 400 m$ . The weakly focusing machine requires extremely large apertures especially to accommodate the energy spread of the beam. In the strongly focusing case, a phase advance of  $60^\circ$  per cell has been chosen.

a)	$\rho = 25 m$	$\ell = 5 m$	b)	$\rho = 400 m$	$\ell = 20 m$
weak foc.	strong foc.		weak foc.	strong foc.	
$\beta = 35 m$	$\hat{\beta} = 17 m$		$\beta = 566 m$	$\hat{\beta} = 69 m$	
$D = 50 m$	$\hat{D} = 5 m$		$D = 800 m$	$\hat{D} = 5 m$	

## 7 ACKNOWLEDGEMENTS

In preparing these lectures and the notes we have profited greatly from the books and articles quoted in the bibliography.

## References

- [1] E.D. Courant and H.S. Snyder: Theory of the Alternating-Gradient Synchrotron, *Annals of Physics* 3, 1958, 1.
- [2] H. Bruck: *Accélérateurs Circulaires de Particules*, Press Universitaires de France, Paris, 1966; english translation: Los Alamos Report LA-TR-72-10
- [3] M. Sands: The Physics of Electron Storage Rings, SLAC report 121, 1970.
- [4] K. Steffen: Basic Course on Accelerator Optics, Proceedings of the 1984 CERN Accelerator School, CERN 85-19, 1985.
- [5] P. Schmüser: Basic Course on Accelerator Optics, Proceedings of the 1986 CERN Accelerator School, CERN 87-10, 1987
- [6] H. Wollnik: Optics of Charged Particles, Academic Press, 1987.
- [7] K. Wille: *Physik der Teilchenbeschleuniger und Synchrotronstrahlungsquellen*, Teubner, Stuttgart, 1992.
- [8] Proceedings of the 1984 CERN Accelerator School, CERN 85-19, 1985.
- [9] R. Brinkmann: Insertions, Proceedings of the 1986 CERN Accelerator School, CERN 87-10, 1987
- [10] K. Steffen: High Energy Beam Optics, Interscience, 1965.
- [11] M. Conte, W.W. MacKay: An Introduction to the Physics of Particle Accelerators, World Scientific, 1991.
- [12] J. Rossbach: Teilchenbeschleuniger, in: *Handbuch der Vakuumelektronik*, Hrsg. Eichmeier/Heynisch, Oldenbourg, 1989, 343-425.
- [13] F. Willeke, G. Ripken: Methods of Beam Optics, DESY 88-114, 1988.
- [14] D. Edwards, M.J. Syphers: An Introduction to the Physics of High Energy Accelerators, Wiley-Interscience, 1993.

# BEAM PHASE SPACE AND EMITTANCE

*J. Buon*

Laboratoire de l'Accélérateur Linéaire, Orsay, France.

## Abstract

The classical and elementary results for canonical phase space, the Liouville theorem and the beam emittance are reviewed. Then, the importance of phase portraits to obtain a geometrical description of motion is emphasized, with examples in accelerator physics. Finally, a statistical point of view is used to define beam emittance, to study its law of approximate conservation, with three particular examples, and to introduce a beam envelope-ellipse and the  $\beta$ -function, emphasizing the statistical features of its properties.

## 1. INTRODUCTION

Beam phase space and beam emittance are topics presented in many textbooks. It is not very necessary to repeat here what can be found elsewhere, particularly in previous CERN Accelerator School courses. It is also assumed that the reader is already familiar with the Hamiltonian formalism of mechanics at least at an elementary level.

To refresh the reader's memory and to guide him, the most important classical and elementary results are reviewed in Section 2. No proofs are given and the reader is referred to the literature. In particular, there are several different proofs of the Liouville theorem of which perhaps the shortest and most direct is that reported by M. Weiss<sup>[1]</sup>. Examples of emittance non-conservation are also only listed, without trying to be exhaustive. After this review, the aim is to develop two particular and important topics : the phase portraits and the statistical concept of emittance.

Section 3 is devoted to the phase portraits : the pictures made by all the possible trajectories in phase space. Their importance lies in the geometrical description of motion that they give. They are particularly useful when the phase space dimensionality is low (2 or 3 at most) and when phase trajectories do not cross. The non-linear pendulum, considered in Section 2 is the simplest example. It is also the case of the uncoupled motion of non-interacting particles submitted to time-independent forces. However, in accelerators particles experience guiding and accelerating forces that vary from place to place. The non-crossing property of the phase trajectories can be restored at the price of increasing the dimensionality by one unit. Fortunately, the Poincaré sections make it possible to reduce the dimensionality later without losing the non-crossing property. The usual turn-by-turn stroboscopic view of the betatron motion in a circular accelerator follows these general lines. It is used here to illustrate the procedure. Finally, some examples of phase portraits in accelerator physics are shown.

In Section 4 a statistical point of view is taken to define and study the beam emittance. It is more satisfactory than the traditional emittance definition using a limiting contour in phase space since the latter suffers from arbitrariness of the contour. Moreover, the statistical definition lends itself more easily to calculations. This definition, initially given by P. Lapostolle<sup>[2]</sup>, is introduced here in the most natural way, as a measure of the spread in phase space of the points representing beam particles. It generalizes in two dimensions the standard deviation that measures the dispersion of points distributed on a line.

With the statistical definition and in the case of a Hamiltonian motion, the conservation of emittance is not a direct consequence of the Liouville theorem as in the usual definition. It is shown that the theorem applies and the emittance is conserved only if the motion is linear. The conclusion is that emittance is conserved, but only approximately, in accelerators where particle motion is made as linear as possible and where other causes of emittance dilution are minimized.

The emittance growth due to either chromaticity or to scattering, and the emittance reduction by stochastic cooling, are studied in an elementary way as three examples of calculation using the statistical definition of emittance.

Finally, to characterize the shape of the spread in phase space, an envelope-ellipse is also defined in a statistical way. It allows one to immediately introduce the  $\beta$ -function as the statistical beam envelope and to avoid any confusion or arbitrariness when using the  $\beta$ -function in unclosed transport lines. The general properties of the  $\beta$ -function are derived to show their statistical features.

## 2. SUMMARY OF CLASSICAL AND ELEMENTARY RESULTS

### 2.1 The canonical phase space

For many physical systems, their evolution after a time  $t$  only depends on their state at that time. This is determined by the values of  $r$  coordinates  $q_i$  ( $i = 1, \dots, r$ ) and their time derivatives  $\dot{q}_i$ , the integer  $r$  being the number of degrees of freedom.

In the absence of friction forces, the equations of motion can be written in the form of Hamilton :

$$\begin{cases} \dot{q}_i = +\frac{\partial H}{\partial p_i} \\ \dot{p}_i = -\frac{\partial H}{\partial q_i} \end{cases} \quad i = 1, \dots, r \quad (1)$$

where the Hamiltonian  $H(q_i, p_i, t)$  is a function of the coordinates  $q_i$ , of the conjugate momenta  $p_i$ , and of time  $t$ . If  $H$  does not explicitly depend on the time  $t$ , the Hamiltonian  $H$  is the total energy of the system that is invariant.

The first set of these equations allows the time derivatives  $\dot{q}_i$  to be expressed in terms of the conjugate momenta  $p_i$ , or vice versa. The second set allows the second-order differential equations of the motion to be derived, the so-called equations of Lagrange.

The canonical phase space is the  $2r$ -dimensional space with the conjugate coordinates  $q_i, p_i$  ( $i = 1, \dots, r$ ). The state of the system at a time  $t$  is represented by a point  $M(t)$  with coordinates  $q_i(t), p_i(t)$  in the canonical phase space. As  $t$  increases, the representative points  $M(t)$  generate a curve, called a phase trajectory.

When the Hamiltonian is time-independent, there is only one phase trajectory originating from any point  $M$  in the phase space, apart from unstable equilibrium points where the derivatives  $\frac{\partial H}{\partial q_i}, \frac{\partial H}{\partial p_i}$  vanish. Different trajectories cannot cross each other. The motion along a closed trajectory is periodic. If the phase space is two-dimensional, all trajectories, originating from points inside the region limited by a closed trajectory, are bounded by this closed trajectory.

A beam of  $N$  particles is a system of  $3N$  degrees of freedom if the internal degrees of freedom (like spin) can be neglected. The  $3N$  coordinates  $q_j$  are the components of the coordinate vectors  $\mathbf{r}_j$  ( $j = 1, \dots, N$ ) of each particle. The canonical phase space is  $6N$ -dimensional.

When the  $N$  particles are identical and without mutual interaction, only the canonical phase space of one particle can be considered. The latter is 6-dimensional. At a time  $t$ , the state of the beam is represented by the set of  $N$  points  $M_j(t)$ , where the point  $M_j$  corresponds to the particle  $j$  ( $j = 1, \dots, N$ ).

The  $N$  particles form a beam only if the  $N$  points are clustered in a relatively small volume of the 6-dimensional phase space. When  $N$  is large, the beam state can be represented by a phase density  $f(\vec{q}, \vec{p}, t)$  such that at time  $t$  the number  $dN$  of representative points in an infinitesimal volume  $d^3pd^3q$  is given by :

$$dN = f(\vec{q}, \vec{p}, t)d^3qd^3p. \quad (2)$$

In general, the phase density  $f(\vec{q}, \vec{p}, t)$  explicitly depends on the time  $t$  as the density does not stay constant at a fixed point  $M(\vec{q}, \vec{p})$ .

The phase density can be considered as a probability distribution in the phase space, resulting from the lack of information on the initial state of the beam. Very often the phase density can be approximated either by a parabolic function, by a Gaussian, or again by a uniform density in an ellipsoid.

Usually, the longitudinal motion along the beam axis is decoupled from the motion in the plane transverse to the beam axis. In that case, the 6-dimensional phase space can be split into a longitudinal phase space (2-dimensional) and a transverse phase space (4-dimensional).

Finally, if the transverse motion can be decomposed into two independent motions along two orthogonal directions, the transverse phase space can also be split into two 2-dimensional phase spaces. This is the ultimate reduction of the phase space dimensionality.

Canonical transformations allow the conjugate variables  $q_i, p_i$  to be changed for new conjugate variables  $Q_i, P_i$ , the canonical phase space  $(q_i, p_i)$  then being mapped into a new phase space  $(Q_i, P_i)$ . One usual canonical transformation in beam dynamics is the change of the longitudinal coordinates  $z, p_z$  for the phase shift  $\delta\varphi$  and the energy increment  $\delta E$ , both relative to a synchronous particle.

Limits are set by quantum mechanics on the knowledge of two conjugate variables  $q_i$  and  $p_i$ . According to the Heisenberg Uncertainty Principle, the product of the uncertainties  $\delta q_i$  and  $\delta p_i$  cannot be made smaller than the Planck constant  $\hbar$  (divided by  $2\pi$ ) :

$$\delta q_i \delta p_i \geq \hbar c = 2 \times 10^{-10} \frac{MeV}{c} \times mm. \quad (3)$$

This limitation can be expressed by saying that the state of a particle is not exactly represented by a point, but by a small uncertainty volume of the order of  $\hbar^3$  in the 6-dimensional phase space.

Nevertheless, the practical accuracy on the knowledge of position and momentum is normally far away from this quantum limit. Classical mechanics can be used to very good approximation to study particle motion, except when there are interactions at the level of atomic size, like the quantum jumps of electrons emitting photons in a static magnetic field (synchrotron radiation).

## 2.2 The Liouville theorem

The evolution of a physical system can also be considered as a mapping of the phase space onto itself, relating the two representative points  $M(t)$  and  $M(t')$  at two different times. When this evolution is governed by a Hamiltonian, the particular form of the equations of motion implies that there are quantities left invariant by the mapping.

In particular, the Liouville theorem states that volumes in phase space are invariant.

By mapping from time  $t$  to time  $t'$ , all the points  $M(t)$  in a finite volume  $V$  are transformed into the points  $M(t')$  that define the transformed volume  $V'$ . From the Liouville theorem, the measures of the volumes  $V$  and  $V'$  are the same. This property applies to the  $6N$ -dimensional canonical phase space of a beam of  $N$  interacting particles, as well as to the 6-dimensional phase space of identical particles without mutual interactions. It is a property of the mapping that results from the Hamiltonian form of the equations of motion and does not depend on the particles themselves. In the 6-dimensional phase space, the invariance of a volume does not depend on how many particles are represented in it. There may be no particles in that volume.

Nevertheless, the hypersurface limiting a finite volume is not invariant in general. Its geometrical form is modified, but in a way such that the enclosed volume is conserved. For instance if the  $N$  points, representing a beam in a 2-dimensional phase space, are enclosed in an ellipse at time  $t$ , then at another time  $t'$ , the  $N$  points are enclosed in a new closed curve, transformed from the ellipse and may be with a different form.

Moreover, volumes in a subspace corresponding to a part of the degrees of freedom are not generally invariant. They are only invariant when these degrees of freedom are uncoupled from the other ones. For instance, when the betatron motions are coupled, areas in horizontal and vertical phase spaces are not conserved.

An immediate consequence of the Liouville theorem and of particle conservation, when beam losses are negligible, is the invariance of the phase density  $f(\vec{q}, \vec{p}, t)$  in the 6-dimensional phase space. More precisely, the density at the point  $M(t)$  is equal to the density at the transformed point  $M(t')$  at another time  $t'$  :

$$f(\vec{q}(t), \vec{p}(t), t) = f(\vec{q}(t'), \vec{p}(t'), t'). \quad (4)$$

In other words, the phase density stays constant when running along a phase trajectory as a representative point does.

The motion of the representative points in the 6-dimensional phase space is analogous to the flow of an incompressible fluid. In this picture, the phase density is analogous to the mass density of the fluid.

However, in general a fluid is homogeneous with a uniform density, while the phase density may not be uniform and often has a gaussian variation in phase space at a given time. On the other hand, there are many examples illustrating the non-invariance of the phase space volumes for non-Hamiltonian systems. For instance, the phase trajectories of a linear damped oscillator are spirals all converging to the origin. Any finite volume shrinks and ultimately vanishes.

### 2.3 The beam emittance

Neglecting mutual interactions and coupling between the three coordinates of a particle, one defines the emittance of each degree of freedom ; horizontal, vertical and longitudinal.

The horizontal (vertical) emittance is usually defined considering the  $x, x' = \frac{dx}{ds}$  phase space, instead of the  $x, p_x$  canonical phase space ( $s$  is the curvilinear abscissa along the reference orbit). One is more interested to know about the slope  $x'$  than the transverse momentum  $p_x$ .

Representative points of particles with the same betatron amplitude, but with different phases, are on the same ellipse at time  $t$ . Homothetic ellipses correspond to different amplitudes.

It is usual to define the emittance  $\epsilon_x$  as the area (optionally divided by  $\pi$ ) of the ellipse containing 95 % of all the particles in its interior (see Fig. 1a). The 95 % proportion is somewhat arbitrary. One could equally use 90 % or 99 %.

The emittance is also the product of the two semi-axes of the ellipse. If the ellipse is upright, its axes coincide with the horizontal and vertical axis. The two semi-axes are half the beam size  $\Delta x$  and the beam divergence  $\Delta x'$  respectively (see Fig. 1b).

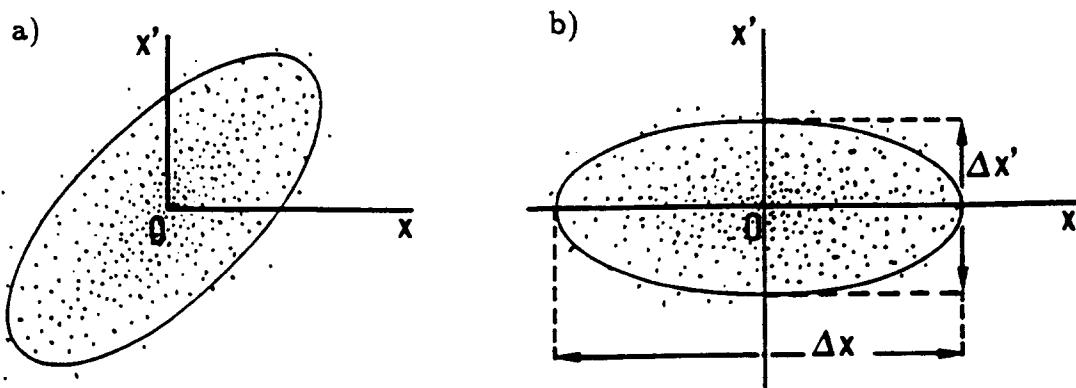


Fig. 1 : A set of points representative of a beam in the  $(x, x')$  phase space  
 a) Tilted emittance ellipse.  
 b) Upright emittance ellipse.

In a beam transport line with negligible longitudinal magnetic field, the potential vector  $\vec{A}$  of the magnetic field is longitudinal. The transverse components of the particle

momentum  $\vec{p} - e\vec{A}$  are equal to the transverse components of the conjugate momentum  $\vec{p}$ . Therefore, the slope  $x'$  is proportional to the conjugate momentum  $p_x$  :

$$p_x = m\beta\gamma cx' \quad (5)$$

where  $m$  is the rest mass of the particles,  $\beta = \frac{v}{c}$  and  $\gamma = 1/\sqrt{1-\beta^2}$  being the usual relativistic quantities attached to the velocity  $v$  of a particle.

In a transport line, or in a ring without acceleration, the energy of a non-radiating particle is constant, and so is  $\beta\gamma$ . The Liouville theorem is also valid in the  $x, x'$  phase space and the emittance  $\epsilon_x$  is conserved. Along the line or the ring, the beam passes through focusing or defocusing lenses. The beam size  $\Delta x$  and the beam divergence  $\Delta x'$  vary, but the emittance stays constant.

In linear, or circular, accelerators, where the particle energy is varied, the emittance is not invariant. Instead one defines the so-called normalized emittance :

$$\epsilon_{xN} = \beta\gamma\epsilon_x \quad (6)$$

that refers to the area in the  $x, p_x$  canonical phase space. The normalized emittance is conserved during acceleration.

As  $\beta\gamma$  increases proportionally to the particle momentum  $p$ , the emittance  $\epsilon$  decreases as  $1/p$ . It is called “adiabatic damping”. An equivalent point of view is to consider that when crossing accelerating cavities, the longitudinal momentum  $p_y$  is increased while  $p_x$  is not changed. Therefore, the slope  $x' = \frac{p_x}{p_y}$  is decreased.

It is worth noting that, due to adiabatic damping, higher-energy accelerators are able to deliver smaller beams to counteract the decrease of the particle interaction cross-section as the energy increases.

An important challenge in accelerator technology is to preserve beam emittance and even to reduce it. In spite of the Liouville theorem, there are many phenomena that may affect the emittance. A list, not necessarily exhaustive, of phenomena causing non-conservation of horizontal (vertical) beam emittance<sup>[3]</sup> includes :

- coupling between degrees of freedom (horizontal-vertical coupling, chromaticity...)
- intrabeam scattering
- beam-beam scattering
- scattering on residual gas
- multiple scattering through a thin foil
- electron cooling
- stochastic cooling
- laser cooling
- synchrotron radiation emission
- filamentation due to non-linearities
- wake fields
- space charge effects.

Some of these processes (coupling, chromaticity, filamentation,...) are Hamiltonian processes that can in principle be compensated to avoid emittance dilution.

The horizontal  $\epsilon_x$  and vertical  $\epsilon_z$  emittances are also related to the beam brightness defined as :

$$B = \frac{I}{\pi^2 \epsilon_x \epsilon_z}. \quad (7)$$

In most applications either emittances or brightness are the main figures of merit of particle beams.

Similarly, the longitudinal emittance measures the product of the energy spread  $\delta E$  by the phase dispersion  $\delta\phi$ . In the linear and smooth approximation of the synchrotron motion, the trajectories of representative points of the particles are ellipses in the  $\delta E, \delta\phi$  plane. The area (divided by  $\pi$ ) of the ellipse containing 95 % of the particles is the longitudinal emittance  $\epsilon_s$ . As the ellipse is now a phase trajectory, the emittance  $\epsilon_s$  is also half the action  $J$  of this trajectory :

$$J = \frac{1}{2\pi} \oint \delta E d(\delta\phi). \quad (8)$$

One can also say that the conservation of the longitudinal emittance is a consequence of the invariance of the action  $J$ . However, the longitudinal Hamiltonian is energy-dependent and varies during acceleration in a synchrotron. Consequently, phase trajectories are not exactly ellipses. Nevertheless, when acceleration is sufficiently slow, ellipses are mapped into ellipses of the same area, according to the adiabatic theorem. The action and the longitudinal emittances are still conserved quantities.

### 3. PHASE PORTRAITS

The usefulness in considering phase space is not just to represent the state of a dynamical system by a point in such a configuration space, and its evolution with time by a phase trajectory. More important is the topological structure of the trajectories in the phase space. All the possible trajectories form a picture, a "phase portrait", characteristic of the dynamical system, with typical topological features. For instance, circular trajectories around a centre, as in small oscillations of a pendulum, reveal a state of stable equilibrium (an elliptic point in phase space). On the contrary, diverging trajectories looking like hyperbolae reveal an unstable equilibrium (a hyperbolic, or saddle, point). Many other topological aspects, like separatrix, limit cycle, spirals, islands, attractors, chaotic attractors,..., reveal other dynamical behaviours.

As was recognized by H. Poincaré, a phase portrait gives a complete geometrical description of all the possible behaviours of a dynamical system. Nevertheless, it could be argued that the analytical solution of the equations of motion for linear dynamical systems is enough to obtain a complete description of their behaviour. With that point of view, phase portraits would only be useful for non-linear systems for which analytical solutions are scarcely available. Particle accelerators are made to be linear systems and phase portraits would not be very useful for studying particle dynamics in accelerators. However, a real dynamical system is only nearly linear and its stability and long-term evolution are essentially determined by its non-linear aspects, even if they are weak. The importance of phase portraits is very general, even for particle dynamics in accelerators.

### 3.1 The non-linear pendulum

The equation of motion of an undamped pendulum is :

$$\frac{d^2\theta}{dt^2} + \omega^2 \sin\theta = 0 \quad (9)$$

with  $\omega = \sqrt{\frac{g}{\ell}}$ ,  $g$  being the gravitation intensity and  $\ell$  the pendulum length (Fig. 2). Due to the  $\sin\theta$  term, that equation is non-linear and consequently the pendulum is a non-linear oscillator. However, it becomes linear at the approximation of small oscillations, for which one obtains the linear equation of a harmonic oscillator :

$$\frac{d^2\theta}{dt^2} + \omega^2\theta = 0. \quad (10)$$

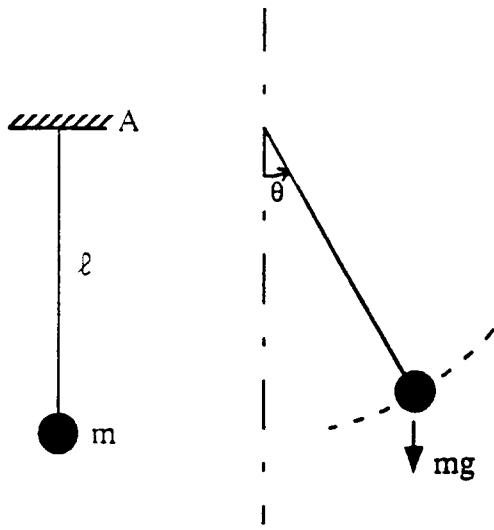


Fig. 2 : A simple and undamped pendulum. The weight of the bob is  $mg$  and the length of the rod is  $\ell$ .

The motion of an undamped pendulum gives the classical and simplest example of phase portrait in a two-dimensional phase space. The motion is governed by a Hamiltonian  $H$  that is identified with the total energy, sum of the kinetic energy  $T$  ( $T = \frac{1}{2m\ell^2} p_\theta^2$ ) and the potential energy  $V$  ( $V = -mg\ell\cos\theta$ ) :

$$H(\theta, p_\theta) = T + V \quad (11)$$

where  $m$  is the mass of the bob.

The Hamilton equations with the conjugate variables  $\theta, p_\theta$  are :

$$\begin{cases} \frac{d\theta}{dt} = +\frac{\partial H}{\partial p_\theta} = \frac{p_\theta}{m\ell^2} \\ \frac{dp_\theta}{dt} = -\frac{\partial H}{\partial \theta} = -mgl \sin\theta. \end{cases} \quad (12)$$

These two equations are equivalent to the second-order differential equation :

$$m\ell^2 \frac{d^2\theta}{dt^2} + mgl \sin\theta = 0. \quad (13)$$

The canonical phase space is two-dimensional with the conjugate variables  $\theta, p_\theta$  as coordinates. The trajectories in the phase space (Fig. 3a) are the curves of constant energy  $E$  :

$$H(\theta, p_\theta) = E. \quad (14)$$

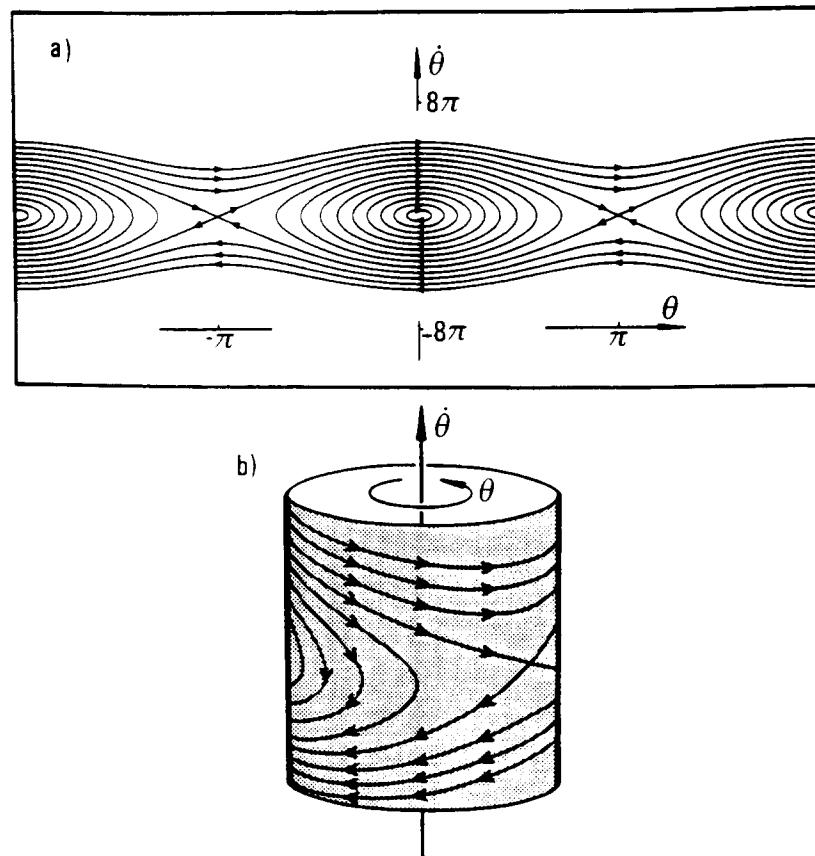


Fig. 3 : Phase trajectories of the non-linear pendulum :  
a) on a plane,  
b) on a cylinder.

The direction of the arrows on these curves are readily obtained by observing that  $\theta$  increases when  $p_\theta$  is positive.

For small-amplitude oscillations, the phase trajectories are ellipses centred on the origin. It is the point of stable equilibrium with the pendulum hanging vertically without swinging.

The second equilibrium position is where the pendulum is balanced vertically on end. It is an unstable equilibrium state for which  $\theta = \pi$  and  $\frac{d\theta}{dt} = 0$ . It appears as a saddle point in phase space, associated with the crossing of a particular phase trajectory, named the separatrix. This is the curve corresponding to the energy of the saddle point. If the pendulum is launched with this energy, it rotates towards the unstable equilibrium position. Its velocity decreases when approaching it such that it would take an infinite time to reach the position.

The wavy lines above and below the separatrix correspond to the whirling motion of the pendulum.

As the pendulum motion is periodic, with a cyclic coordinate  $\theta$  (the state of the pendulum is unchanged by adding  $2\pi$ ), one can also use a cylindrical phase space as illustrated in Fig. 3b.

### 3.2 The non-crossing trajectory property

The phase portrait is especially clear when phase trajectories do not cross, as illustrated by the example of the pendulum. It is an advantage to have the “non-crossing” condition.

To also include the case of a separatrix crossing itself at a saddle point, one only requires that there is no crossing in finite time. A sufficient condition is that the Hamiltonian of the system should not explicitly depend on time. Therefore, the Hamilton equations show that the tangent to the phase trajectory at any point is uniquely determined (if not an equilibrium position). From this one can infer the uniqueness of the phase trajectory originating from that point.

The topological consequences of this non-crossing property are best illustrated in the case of a two-dimensional phase space. If a phase trajectory is a closed loop, the corresponding motion is periodic. This trajectory divides the plane into two regions, the interior and the exterior. Any trajectory, starting from inside, stays inside for ever and is then bounded.

When the Hamiltonian is time-dependent, one can nevertheless satisfy the non-crossing property at the price of increasing the phase space dimensionality. It is exactly what happens for betatron motion and it can be taken as an example of the procedure.

In the linear approximation, the betatron motion along one direction  $x$  is governed by an equation of Hill's type :

$$\frac{d^2x}{dt^2} + \omega^2(t)x = 0. \quad (15)$$

The angular frequency  $\omega$  depends on time  $t$  as the focusing strength varies along the ring. The Hamiltonian  $H$  is :

$$H = \frac{p_x^2}{2m} + \frac{1}{2} m \omega^2(t) x^2. \quad (16)$$

To the canonical variables  $x$  and  $p_x$ , one adds a third variable  $s = ct$  that can be interpreted as the abscissa of a reference particle running with velocity  $v$  along a reference orbit. The system of the two Hamilton equations :

$$\begin{aligned} \frac{dx}{dt} &= \frac{p_x}{m} \\ \frac{dp_x}{dt} &= -m\omega^2(s)x \end{aligned} \quad (17)$$

becomes a system of three equations, with the third equation :

$$\frac{ds}{dt} = v. \quad (18)$$

The phase space is now three-dimensional with the coordinates  $x, p_x$  and  $s$ . There is only one phase trajectory originating from any point  $(x, p_x, s)$ . Figure 4 shows an example of a betatron phase trajectory in the  $(x, p_x, s)$  phase space.

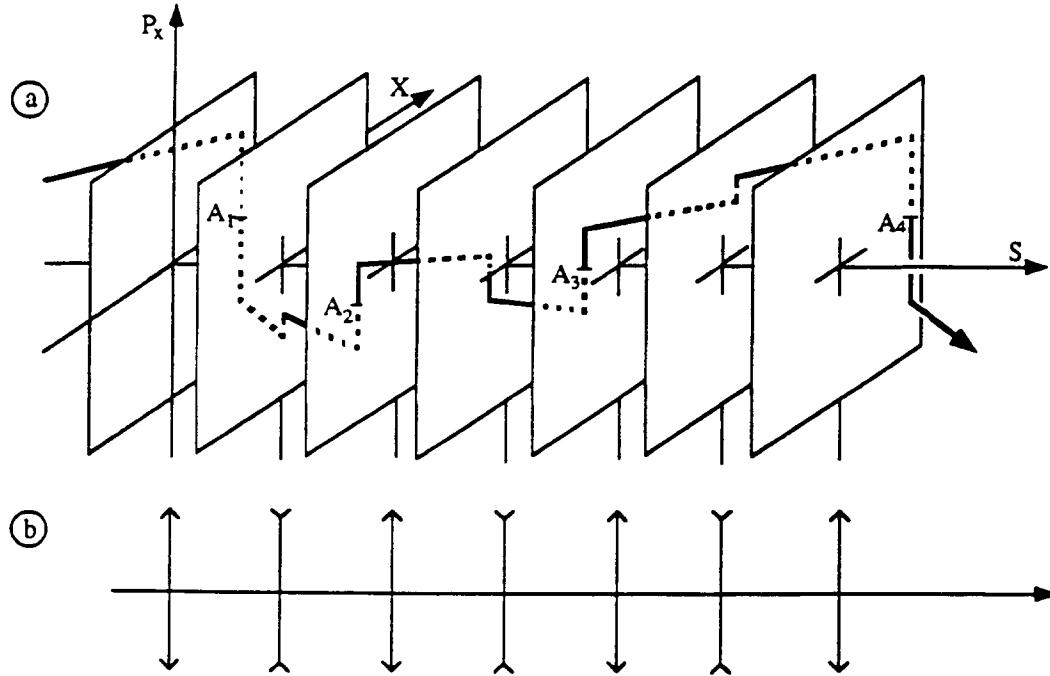


Fig. 4 : A betatron phase trajectory in the  $(x, p_x, s)$  phase space (a), corresponding to a periodic FODO transport line in the thin-lens approximation (b). Only the crossing  $A_1, A_2, A_3, A_4$  with the transverse planes at the position of the focusing lenses are indicated. The betatron phase advance is  $120^\circ$  per period.

### 3.3 The Poincaré section

The phase portrait gives a geometrical and global view of motion. However, this view becomes difficult to perceive when the phase space dimensionality is larger than 2 or 3.

Nevertheless, the topological structure can still be revealed by only looking at the intersection of the phase trajectories with a hypersurface in phase space. Such a hypersurface is called a Poincaré section.

Again, one can illustrate the procedure with the example of the betatron motion, here in the periodic lattice of a ring. The angular frequency  $\omega(t)$  is periodic with the period  $T = L/v$  ( $L$  is the lattice period).

In the three-dimensional  $(x, p_x, s)$  phase space one looks at the intersection of any phase trajectory with the planes  $x, p_x$  whenever  $s$  is a multiple of the lattice period  $L$  (see Fig. 5). These intersections give a series of  $x, p_x$  dots. It can be considered as a stroboscopic technique where a snapshot of the phase trajectory is taken at each period  $L$ .

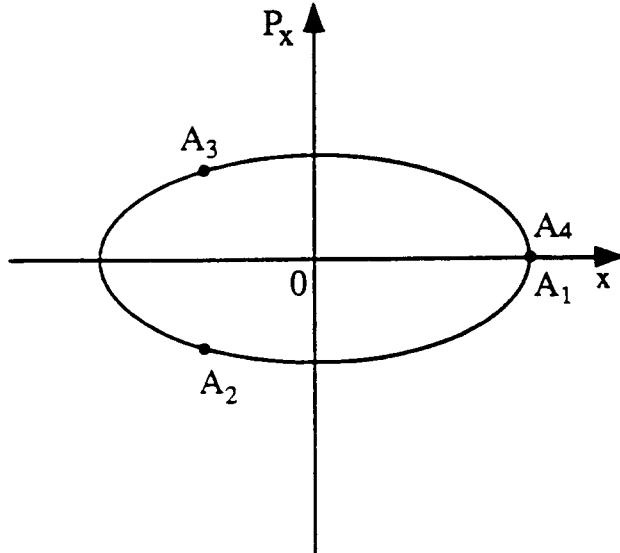


Fig. 5 : Intersection points  $A_1, A_2, A_3, A_4$  of the phase trajectory shown on Fig. 4.  $A_1$  and  $A_4$  coincide as the tune is  $1/3$ . The ellipse passing through the points is also shown.

The series of  $x, p_x$  dots is a mapping of the  $(x, p_x)$  plane into itself, called a Poincaré mapping.

For a linear lattice, this Poincaré mapping is a linear mapping. Moreover, from the Floquet theorem it results that the dots are located on a common ellipse, assuming that the motion is stable.

If the phase trajectories have the same period as the lattice, any dot is mapped into itself and the series is reduced to a single dot. It corresponds to an integer betatron tune.

If the phase trajectories are closed after  $n$  periods, the dot series reduces to  $n$  points. The betatron tune is a rational  $m/n$ .

Due to the periodicity of the lattice, the Poincaré mapping is invariant by a translation of  $L$ . It means that the Poincaré mapping is time-independent and satisfies the non-crossing property. The ellipses in the  $(x, p_x)$  plane do not cross and one has the phase portrait of a linear oscillator.

Similarly, the finite difference equations of the synchrotron motion between RF cavities also define a Poincaré mapping.

### 3.4 Examples of phase portraits in accelerator physics

Three examples taken from previous CAS lectures will be briefly mentioned, the reader being referred to the CAS proceedings mentioned for further details.

Resonant extraction uses the action of a sextupole non-linearity near a third-integer resonance<sup>[4,5]</sup>. The time-dependence of the Hamiltonian is suppressed by the use of a rotating reference frame. Figure 6 shows the phase portrait in the  $J, \varphi$  canonical phase space.

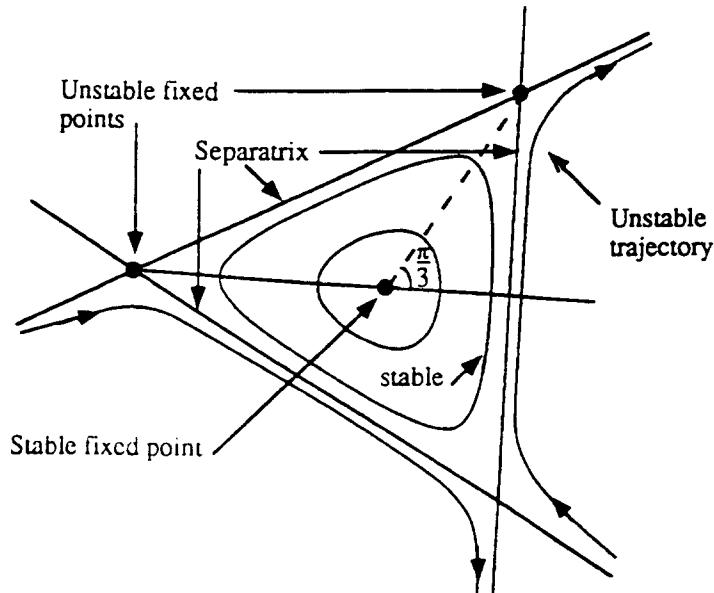


Fig. 6 : Phase portrait of a resonant extraction<sup>[4,5]</sup>

The phase portrait (Fig. 7) of the synchrotron motion depends on the RF phase  $\phi_s$  of the synchronous particle<sup>[6,7,8]</sup>. The phase space domain inside the separatrix is the RF bucket.

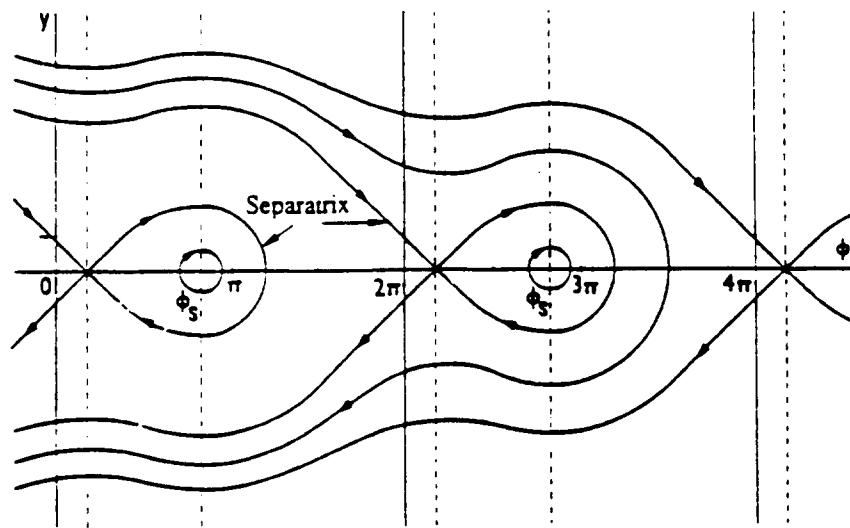


Fig. 7 : Phase portrait of synchrotron motion<sup>[6,7,8]</sup> ( $\frac{\pi}{2} < \phi_s < \pi$ )

The betatron motion, driven by the non-linear beam-beam force, is commonly studied by the technique of the Poincaré sections and mapping. Two-dimensional Poincaré sections show peculiar phase portraits<sup>[9]</sup>, with eventually stochastic domains (Fig. 8).

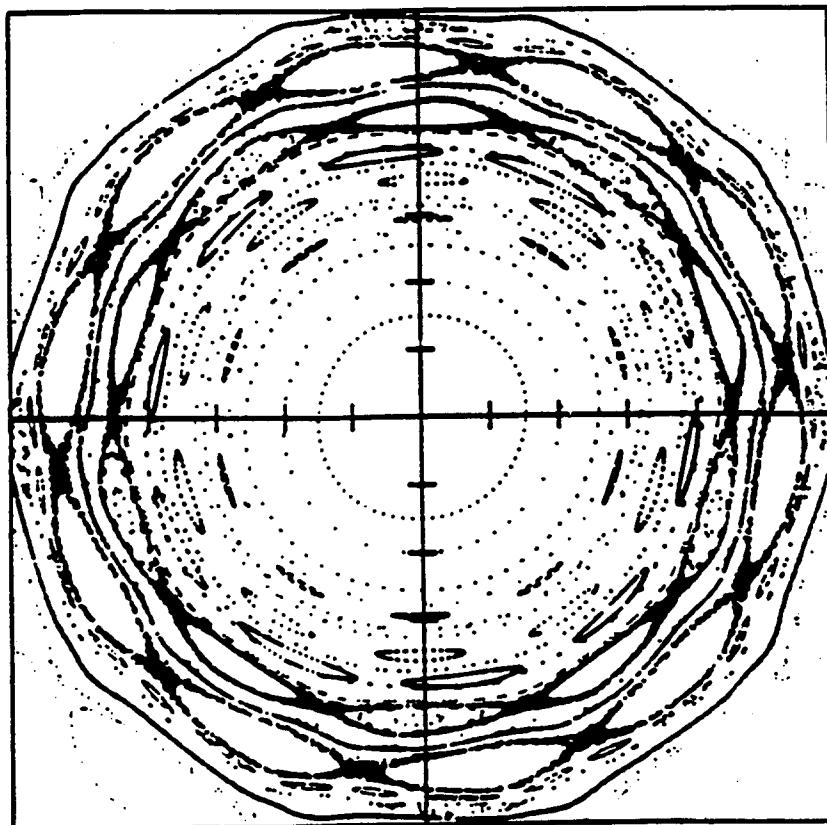


Fig. 8 : Phase portrait of betatron motion perturbed by a beam-beam force<sup>[9]</sup>

## 4. BEAM EMITTANCE : A STATISTICAL POINT OF VIEW

Considering a beam as a statistical set of points in a two-dimensional phase space, the beam emittance is a measure of their spread in that plane (see Fig. 1). The beam emittance is similar to the standard deviation  $\sigma$  that measures the dispersion of a set of points on a line :

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2} \quad (19)$$

where  $\bar{x}$  is the average of the abscissa  $x_i$  of each point. Similarly, the beam emittance should be a “root mean square” quantity, i.e. a function of the second-order moments of the point distribution.

Within the statistical point of view, the emittance is not related to any contour limiting the area occupied by the points. The usual definition of emittance as the area of a limiting contour suffers not only from the arbitrariness of the chosen contour, but also from its distortion when time is running. It may become so twisted and bent that its area is no more representative of the spread of the particles. Higher-order moments of the point distribution could also be introduced to parametrize other aspects, like skewness, but the emittance is the main parameter of the distribution.

### 4.1 The statistical definition

A distribution of points  $M(x, x')$  in phase space can be translated and rotated without changing its spread. One is free to choose an origin  $O$  and two coordinate axes  $X$  and  $X'$ . It is natural to take the origin at the barycentre of the points, such that the averages  $\bar{x}$  and  $\bar{x}'$  of their coordinates  $x_i, x'_i$  are null.

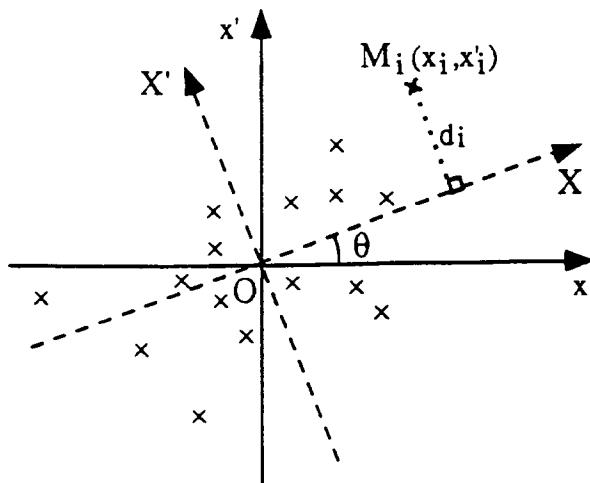


Fig. 9 : The distance  $d_i$  to an axis with polar angle  $\theta$

It is also convenient to orientate the axes  $X$  and  $X'$  to minimize (maximize) the sum of their squared distances to these axes.

If  $\theta$  is the polar angle of such an axis, the distance  $d_i$  of a point  $M_i(x_i, x'_i)$  to this axis is (see Fig. 9) :

$$d_i = |x'_i \cos \theta - x_i \sin \theta|. \quad (20)$$

The mean square distance  $\sigma_{X'}^2 = \frac{1}{N} \sum_{i=1}^N d_i^2$  is minimized with respect to the angle  $\theta$ .

The solution of  $\frac{\partial}{\partial \theta} \sigma_{X'}^2 = 0$  gives :

$$\tan 2\theta = \frac{2 \overline{xx'}}{\overline{x^2} - \overline{x'^2}} \quad (21)$$

where :

$$\overline{x^2} = \frac{1}{N} \sum_{i=1}^N x_i^2, \quad \overline{x'^2} = \frac{1}{N} \sum_{i=1}^N x'^2_i \quad \text{and} \quad \overline{xx'} = \frac{1}{N} \sum_{i=1}^N x_i x'_i \quad (22)$$

are the central second-order moments of the point distribution since  $\bar{x} = \bar{x}' = 0$ . It can be seen that the quantity to minimize is not the same as the one used to find the regression line. In general, the axis will not coincide with the latter. In fact, two solutions  $\theta$  and  $\theta + \frac{\pi}{2}$  giving two orthogonal axes  $X$  and  $X'$  have been found, with two extremum values  $\sigma_X^2$  and  $\sigma_{X'}^2$  :

$$\sigma_{X, X'}^2 = \frac{1}{2} \left( \overline{x^2} + \overline{x'^2} \pm \frac{2 \overline{xx'}}{\sin 2\theta} \right). \quad (23)$$

$\sigma_X^2$  (resp.  $\sigma_{X'}^2$ ) measures the dispersion of the points in the direction of the  $X$  axis (resp.  $X'$ ). It is then natural to define the emittance  $\epsilon$ , i.e. the spread of the distribution, as :

$$\epsilon = 2\sigma_X \cdot 2\sigma_{X'} \quad (24)$$

The expression so obtained :

$$\epsilon = 4 \sqrt{\overline{x^2} \cdot \overline{x'^2} - (\overline{xx'})^2}. \quad (25)$$

was first given by P. Lapostolle<sup>[2]</sup>, and is sometimes called either “effective emittance” or “r.m.s. emittance”. The factor 4 in front of the square root is optional. The expression in the square root is also the determinant of the covariance matrix :

$$\epsilon = 4 \sqrt{\det \begin{pmatrix} \overline{x^2} & \overline{xx'} \\ \overline{xx'} & \overline{x'^2} \end{pmatrix}}. \quad (26)$$

With the use of the standard deviations  $\sigma_x$ ,  $\sigma_{x'}$  and the correlation coefficient  $r$  :

$$r = \frac{\overline{xx'}}{\sqrt{\overline{x^2} \cdot \overline{x'^2}}},$$

the expression of the emittance can also be written as :

$$\epsilon = 4 \sigma_x \sigma_{x'} \sqrt{1 - r^2}, \quad (28)$$

and the latter expression is still valid when the mean values  $\bar{x}$  and  $\bar{x'}$  do not vanish (i.e. when the origin is arbitrary). The absolute value of the correlation coefficient is always less than (or equal to) one and the expression in the square root of (25) is semi-definite positive.

**Exercise :** Prove the preceding statement by expressing that the mean square of  $x + \lambda x'$  can never be negative whatever  $\lambda$ .

Now, a straightforward algebraic manipulation of the equation (25) produces :

$$\epsilon = \frac{2\sqrt{2}}{N} \sqrt{\sum_{i=1}^N \sum_{j=1}^N (x_i x'_j - x_j x'_i)^2}. \quad (29)$$

The expression in the square root is just the sum of the squared areas  $A_{ij}^2$  of all the  $N(N - 1)$  triangles formed by any couple of points  $M_i, M_j$  and the origin O (ignoring a factor 2) :

$$\epsilon = \frac{4}{N} \sqrt{2 \sum_{i=1}^N \sum_{j=1}^N A_{ij}^2} \quad (30)$$

(Caution : each  $M_i, M_j$  couple enters twice in the double sum).

Thus the emittance can be considered as a statistical mean area : the r.m.s. area of the triangles  $OM_i M_j$ . One could have taken that r.m.s. area to measure the spread of points and to define emittance as well. Here again, this statistical mean area is not the area of a limiting contour, but the measure of the spread of the points around their barycentre. When their spread increases, the r.m.s. triangle area increases.

This statistical definition of the emittance leads to some differences with the usual definition with a contour. For instance, both definitions give a null emittance to a distribution on a straight line in the phase space. However, if the line has a curved shape, the usual emittance is still null (using a non-elliptic contour in this case), while the statistical one is not (Fig. 10).

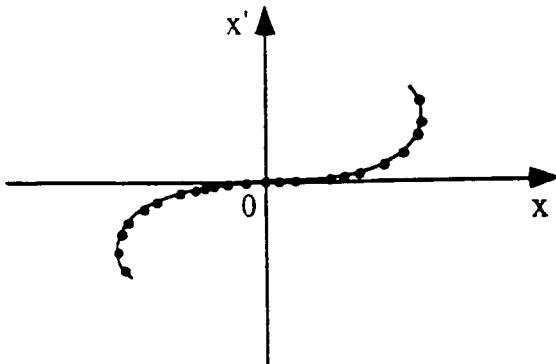


Fig. 10 : Distribution of points on a curved line

This example also shows that the emittance is not unambiguously related to the entropy. According to the Boltzmann definition, the entropy of such a set of points is the logarithm of the area that they occupy in the two-dimensional phase space. Again in this example, the area is null for a curved line, while the statistical emittance is not. It is worthwhile to remark that when the occupied area becomes so twisted and bent, as in the filamentation process, that area is apparently increased and the entropy increases as the emittance does.

Care is also needed with the tail of a point distribution when considering the statistical emittance  $\epsilon$ , as well as the standard deviation. Points at very large distance from the barycentre give large contribution to  $\epsilon$ . If there is a long but small tail, the statistical emittance overestimates the spread of the distribution.

In the case of a long tail, one can also extend to two dimensions the usual alternative to measure the spread of a one-dimensional distribution. Instead of the standard deviation, the spread is better represented by the full width at half maximum (FWHM) of the distribution. Figure 11a shows a one-dimensional Gaussian distribution with a standard deviation 1 on which is superimposed a 10 % Gaussian tail with a standard deviation 10. The tail increases the overall standard deviation by 29 %, while the FWHM is only 7 % larger. Similarly, the emittance can be represented as the area limited by the contour at half-maximum. Figure 11b shows a bi-dimensional Gaussian distribution with standard deviations  $\sigma_x = 2$  and  $\sigma_{x'} = 1$ . A 10 % Gaussian tail, with standard deviations  $\sigma_x = \sigma_{x'} = 10$ , is superimposed.

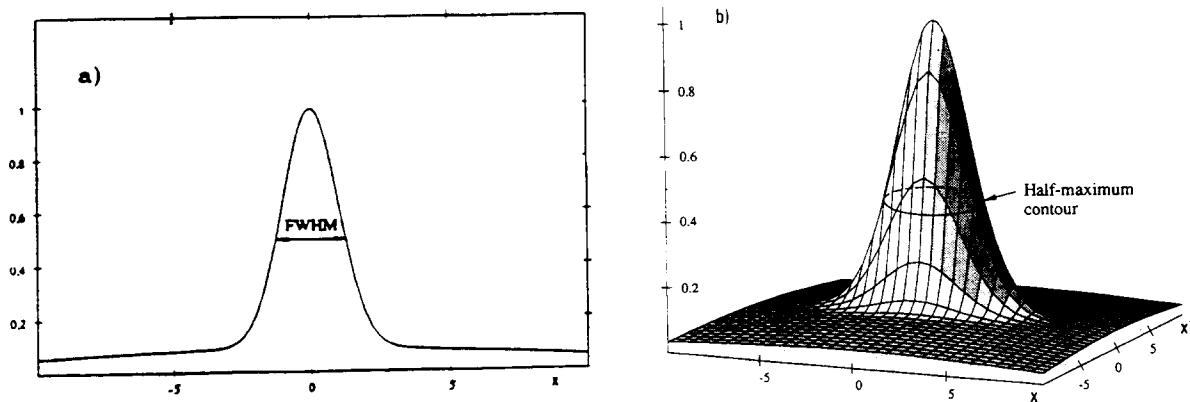


Fig. 11 : a) Full width at half-maximum of a one-dimensional distribution,  
b) Contour at half-maximum of a two-dimensional distribution.

- Exercises :
- 1) Show that the emittance of three points at the tips of an equilateral triangle, inscribed in a circle of radius  $R$ , is :  $\epsilon = 2R^2$ .
  - 2) Show that the emittance of a uniform distribution in a circle of radius  $R$  is  $\epsilon = R^2$ .
  - 3) A distribution of points has a Gaussian density :

$$f(x, x') = \frac{1}{2\pi\sigma_x\sigma_{x'}} e^{-\frac{x^2}{2\sigma_x^2} - \frac{x'^2}{2\sigma_{x'}^2}}.$$

Show that the probability  $p$  that a point lies outside the ellipse :

$$\frac{x^2}{2\sigma_x^2} + \frac{x'^2}{2\sigma_{x'}^2} = c,$$

is  $p = e^{-c}$ . Show that the emittance  $\epsilon_0$ , traditionally defined as the area (divided by  $\pi$ ) of the ellipse containing 95 % of the points, is  $\epsilon_0 = 6 \sigma_x \sigma_{x'}$  and that the ratio of the r.m.s. emittance  $\epsilon$  and of the traditional emittance  $\epsilon_0$  is  $\frac{\epsilon}{\epsilon_0} = \frac{2}{3}$ .

#### 4.2 Emittance conservation

Consider now the time evolution of the beam emittance  $\epsilon$ . The particle motion is assumed to be Hamiltonian, not significantly coupled and in general non-linear :

$$H(x, x') = \frac{x'^2}{2} + f(x). \quad (31)$$

The Hamilton equations of motion, with respect to the independent variable  $s$ , are :

$$\begin{aligned} \frac{dx}{ds} &= x' \\ \frac{dx'}{ds} &= -\frac{\partial H}{\partial x}. \end{aligned} \quad (32)$$

The equations of motion of the barycentre  $(\bar{x}, \bar{x}')$  are :

$$\begin{aligned} \frac{d\bar{x}}{ds} &= \bar{x}' \\ \frac{d\bar{x}'}{ds} &= -\frac{\partial \bar{H}}{\partial x}. \end{aligned} \quad (33)$$

If the motion is linear,  $\frac{\partial \bar{H}}{\partial x}$  is proportional to  $\bar{x}$ . Then the barycentre stays at rest ( $\frac{d\bar{x}}{ds} = \frac{d\bar{x}'}{ds} = 0$ ) if it coincides with the origin ( $\bar{x} = \bar{x}' = 0$ ) at some time.

However, if the motion is non-linear, the time derivative of  $\bar{x}'$  involves higher-order moments of the coordinate  $x$ , and generally does not vanish. The barycentre moves in the phase space.

Moreover,  $\frac{\partial \bar{H}}{\partial x}$  is not in general equal to  $\frac{\partial H}{\partial x}(\bar{x})$ . It means that in the non-linear case, the barycentre does not generally move according to the same law as the representative points. Even if this was so, the straight lines joining a couple of points  $M_i, M_j$  and the barycentre  $O$  are not mapped into straight lines, and the triangle  $OM_iM_j$  is not mapped into another triangle, except in the linear case. Therefore, the Liouville theorem does not imply that the area  $A_{ij}$  of the triangles  $OM_iM_j$  is conserved.

The emittance here defined as a sum of areas  $A_{ij}$  is not conserved except when the motion is linear. That is in contrast with the conservation of the traditionally defined emittance due to the Liouville theorem.

To obtain this result analytically, one calculates the derivative  $\frac{d\epsilon}{ds}$  of the emittance. Here, one cannot in general assume  $\bar{x} = \bar{x}' = 0$ . One must replace  $\bar{x}^2$ ,  $\bar{x}'^2$  and  $\bar{x}\bar{x}'$  by the central second-order moments  $\mu_{20}$ ,  $\mu_{02}$  and  $\mu_{11}$  respectively in the expression of the emittance :

$$\epsilon = 4 \sqrt{\mu_{20} \mu_{02} - \mu_{11}^2}, \quad (34)$$

where :

$$\begin{aligned} \mu_{20} &= \sigma_x^2 = \overline{(x - \bar{x})^2} = \bar{x}^2 - \bar{x}^2 \\ \mu_{02} &= \sigma_{x'}^2 = \overline{(x' - \bar{x}')^2} = \bar{x}'^2 - \bar{x}'^2 \\ \mu_{11} &= r\sigma_x\sigma_{x'} = \overline{(x - \bar{x})(x' - \bar{x}')} = \bar{x}\bar{x}' - \bar{x} \cdot \bar{x}'. \end{aligned} \quad (35)$$

A straightforward calculation gives the derivatives of the moments :

$$\begin{aligned} \frac{d\mu_{20}}{ds} &= 2\mu_{11} \\ \frac{d\mu_{02}}{ds} &= -2\mu_{10H} \\ \frac{d\mu_{11}}{ds} &= \mu_{02} - \mu_{10H}, \end{aligned} \quad (36)$$

where :

$$\begin{aligned} \mu_{10H} &= \bar{x} \frac{\overline{\partial H}}{\partial x} - \bar{x} \cdot \frac{\overline{\partial H}}{\partial x} \\ \mu_{01H} &= \bar{x}' \frac{\overline{\partial H}}{\partial x} - \bar{x}' \cdot \frac{\overline{\partial H}}{\partial x}. \end{aligned} \quad (37)$$

With these relations, the derivative  $\frac{d\epsilon}{ds}$  of the emittance is :

$$\frac{d\epsilon}{ds} = \frac{16}{\epsilon} (\mu_{11} \mu_{10H} - \mu_{20} \mu_{01H}). \quad (38)$$

This equation has been derived in a different way by I. Hofmann<sup>[10]</sup>.

In the case of a linear motion :

$$\frac{\partial H}{\partial x} = ax, \quad \frac{\overline{\partial H}}{\partial x} = a\bar{x}, \quad (39)$$

one has :

$$\mu_{10H} = a\mu_{20} \quad \text{and} \quad \mu_{01H} = a\mu_{11}.$$

The derivative  $\frac{d\epsilon}{ds}$  vanishes and the emittance is conserved.

When the motion is non-linear,  $\frac{d\epsilon}{ds}$  cannot vanish for all  $s$ , and the emittance is not conserved. In practical situations, one generally observes an increase of the emittance through the process of filamentation. However, non-linearities are harmful to the stability of motion. They are maintained at the smallest possible level in particle accelerators. Moreover, other causes of emittance dilution (coupling, scattering, radiation,...) are also minimized.

It can be concluded that in accelerators the beam emittance is nearly conserved, but not exactly.

**Exercise :** Show that, assuming a single even multipole term in the Hamiltonian and neglecting the coupling with the other transverse coordinate  $z$  :

$$H = \frac{x'^2}{2} + a_{2n} \frac{x^{2n}}{2n}, \quad (n > 1)$$

the barycentre stays at rest if it coincides with the origin 0 at an initial time and if the point distribution is symmetrical with respect to the origin. Prove without any calculation that the emittance is not conserved.

### 4.3 Emittance growth due to chromaticity

Chromaticity is the variation of the focusing strength of magnetic lenses with particle momentum. It is an example of coupling between momentum and betatron motion that leads to an emittance growth. Let one only consider the simple case of a thin quadrupole with an integrated focusing strength :

$$K(\epsilon) = K(0)(1 - \epsilon),$$

where the quantity :

$$1 - \epsilon = \frac{1}{1 + \delta},$$

measures its variation with the momentum deviation  $\delta = \frac{\Delta P}{P}$ .

From the transformation law through the quadrupole :

$$\begin{pmatrix} x_2 \\ x'_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ k(\epsilon) & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x'_1 \end{pmatrix},$$

one calculates the variation of second-order moments :

$$\begin{aligned}\Delta \overline{x^2} &= 0 \\ \Delta \overline{x'^2} &= K(0) \overline{\delta^2} (2\overline{xx'} + 3K(0) \overline{x^2}) \\ \Delta \overline{xx'} &= K(0) \overline{\delta^2} \overline{x^2},\end{aligned}$$

and the variation of the emittance :

$$\frac{\Delta \epsilon}{\epsilon} = \frac{1}{2} \xi^2 \overline{\delta^2},$$

where one has introduced the chromaticity of the quadrupole  $\xi = K(0)\beta$ .

That chromatic effect introduces a correlation between  $x$ ,  $x'$  and  $\delta$  that can be compensated by a sextupole located in a dispersive region. That is the principle of the chromatic correction.

#### 4.4 Emittance growth by scattering

The statistical definition of beam emittance also allows the emittance growth by scattering to be calculated very simply.

Consider a beam passing through a thin foil where particles are scattered at random. Considering the horizontal direction  $x$  only, when a particle is scattered, its position  $x$  is not changed, but its horizontal slope is changed from  $x'$  to  $x' + \delta x'$ . The change  $\delta x'$  is independent of  $x$  and  $x'$  and averages to zero. The change of  $\overline{x'^2}$  and  $\overline{xx'}$  are :

$$\begin{aligned}\Delta \overline{x'^2} &= \overline{(x' + \delta x')^2} - \overline{x'^2} = \overline{\delta x'^2} \\ \Delta \overline{xx'} &= \overline{x(x' + \delta x')} - \overline{xx'} = 0.\end{aligned}\tag{40}$$

Finally, according to the statistical definition, the increase of the emittance is simply given by :

$$\Delta \epsilon^2 = 16 \overline{x^2 \delta p^2}.\tag{41}$$

A similar use of the statistical definition for studying emittance growth by radiation fluctuations can be found in a paper by M. Sands<sup>[11]</sup>.

#### 4.5 Stochastic cooling

The statistical definition of beam emittance is particularly useful in understanding the principle of stochastic cooling which aims to reduce the statistical spread of particles in phase space. It is based on a non-Hamiltonian process and one cannot claim any violation of the Liouville theorem in spite of an apparent decrease of a phase space area.

The principle of stochastic cooling can be described as follows<sup>[12]</sup> : A pick-up electrode makes it possible to select a small sample of  $n$  particles among the  $N$  particles of a beam ( $n \ll N$ ), and to measure their mean displacement  $\bar{x}$ , i.e. the  $x$ -barycentre of the sample. The latter may be finite, even if the  $x$ -barycentre of the beam is vanishing.

The signal on the electrode is proportional to  $\bar{x}$  and is used to apply a correction signal to another electrode such that the  $x$ -barycentre of the sample is reduced, eventually down to zero if the correction is complete. This procedure is repeated many times on different samples of particles from the same beam, so reducing the beam emittance.

In order to understand the mechanism of stochastic cooling, the simplest way is to consider samples of only one particle ( $n = 1$ ) whose position  $x_0$  becomes zero after correction. This correction modifies the second-order momenta of the  $N$  particles :

$$\bar{x^2} \rightarrow \bar{x^2} - \frac{x_0^2}{N}, \quad \bar{x'^2} \rightarrow \bar{x'^2} \text{ and } \bar{xx'} \rightarrow \bar{xx'} - \frac{x_0 x'_0}{N}. \quad (42)$$

The change of the emittance is :

$$\begin{aligned} \Delta\epsilon^2 &= 16 \left[ \left( \bar{x^2} - \frac{x_0^2}{N} \right) \bar{x'^2} - \left( \bar{xx'} - \frac{x_0 x'_0}{N} \right)^2 - \bar{x^2} \cdot \bar{x'^2} + \bar{xx'}^2 \right] \\ \Delta\epsilon^2 &= -\frac{16}{N} \left[ x_0^2 \bar{x'^2} - 2 x_0 x'_0 \bar{xx'} + \frac{x_0^2 x'_0^2}{N} \right]. \end{aligned} \quad (43)$$

As the correction procedure is repeated on many samples, the final change of the emittance is expected to be given by the preceding expression averaged over all the possible  $x_0, p_0$  values. One has just to replace  $x_0^2$  and  $x_0 p_0$  by their respective averaged values  $\bar{x^2}$  and  $\bar{xp}$ .

Finally, neglecting the  $\frac{1}{N}$  term in the bracket, one obtains :

$$\Delta\epsilon^2 = -\frac{16}{N} [\bar{x^2} \cdot \bar{x'^2} - 2 \bar{xx'}^2] \quad (44)$$

or :

$$\Delta\epsilon^2 = -\frac{16}{N} \bar{x^2} \cdot \bar{x'^2} (1 - 2r^2). \quad (45)$$

The emittance is reduced only if  $r < \sqrt{\frac{1}{2}}$ . It is easily seen that the emittance would be increased for larger values of  $r$ , in particular when  $x$  and  $p$  are linearly dependent ( $r = 1$ ). The correction is also most efficient when  $x$  and  $p$  are not correlated ( $r = 0$ ). In fact, the correction is incomplete and  $x_0$  is only replaced by  $(1 - g)x_0$  where  $g$  is the gain ( $0 < g < 1$ ). The final expected change  $\Delta\epsilon^2$  becomes (the proof is left as an exercise) :

$$\Delta\epsilon^2 = -\frac{16}{N} \bar{x^2} \cdot \bar{x'^2} g(2 - g - 2r^2). \quad (46)$$

In practice, the gain is small and the emittance is reduced, except if  $r$  is very close to 1.

## 4.6 The beam envelope-ellipse and the $\beta$ -function

The emittance, that measures the spread of particles in phase space, does not give all the information that is contained in the second-order moments. In particular, it is useful to know the geometrical shape of the particle spread in phase space. It can be made more explicit by defining an ellipse with parameters involving the second-order moments of their distribution. It will allow to introduce the  $\beta$ -function in a statistical way. The axes of that envelope-ellipse will be the two axes  $X$  and  $X'$  that minimize (resp. maximize) the mean squared distances of the points to the axes. These axes are rotated by an angle  $\theta$  with respect to the axes  $x$  and  $x'$  (see section 4.1 and figure 9). The two semi-axes  $a$  and  $b$  of the envelope-ellipse are made proportional to the r.m.s. values of these distances to the two axes  $X$ ,  $X'$  and fixed to :

$$a = 2\sigma_X, \quad b = 2\sigma_{X'},$$

such that the area  $A$  of the envelope-ellipse is just  $\pi$  times the emittance  $\epsilon$  :

$$A = \pi\epsilon. \quad (47)$$

With respect to the axes  $X$  and  $X'$  the equation of the envelope-ellipse is :

$$\frac{X^2}{4\sigma_X^2} + \frac{X'^2}{4\sigma_{X'}^2} = 1.$$

By an inverse rotation of angle  $-\theta$  one obtains its equation with respect to the  $x$  and  $x'$  axes :

$$x^2\sigma_x^2 - 2xx'r\sigma_x\sigma_{x'} + x'^2\sigma_{x'}^2 = \frac{\epsilon^2}{4}, \quad (48)$$

using the standard deviations  $\sigma_x$ ,  $\sigma_{x'}$  and the correlation coefficient  $r$ . That equation can also be written in the conventional form :

$$\gamma x^2 + 2\alpha xx' + \beta x'^2 = \epsilon, \quad (49)$$

defining the  $\alpha$ ,  $\beta$ ,  $\gamma$  parameters such that :

$$\begin{aligned} 2\sigma_x &= 2\sqrt{x^2} = \sqrt{\beta}\epsilon \\ 2\sigma_{x'} &= 2\sqrt{x'^2} = \sqrt{\gamma}\epsilon \\ 4r\sigma_x\sigma_{x'} &= 4\overline{xx'} = -\alpha\epsilon. \end{aligned} \quad (50)$$

- General remarks on the envelope-ellipse and the parameters  $\alpha$ ,  $\beta$ ,  $\gamma$  :

- i) for a fixed emittance, the parameter  $\sqrt{\beta}$  is proportional to the r.m.s. beam dimension  $\sigma_x$ . It represents the *beam-envelope* per unit emittance. Similarly,  $\sqrt{\gamma}$  is proportional to the r.m.s. beam divergence and  $\alpha$  is proportional to the correlation between  $x$  and  $x'$ . From Eqs. (28) and (50) one obtains :

$$\alpha = -\frac{r}{\sqrt{1-r^2}}.$$

The factor 2, like the factor 4 in the definition of the r.m.s. emittance, is optional. Both are chosen such that a large majority of the points are inside the envelope-ellipse.

Exercise : Show that the percentage of points inside the envelope-ellipse is 86.5 % in the case of a gaussian distribution (use the first result of the exercise 3 in section 4.1).

- ii) The statistical  $\alpha, \beta, \gamma$  parameters can *always* be defined, while  $\alpha, \beta, \gamma$  traditionally defined in a periodic lattice cannot be defined if the lattice is not linear (one only calculates the  $\beta$ -function for the linear part of the lattice, without considering the presence of non-linear fields like in sextupoles for instance).
- iii) When an emittance dilution occurs, in general there is also a perturbation of the  $\alpha, \beta, \gamma$  parameters. The changes of  $\epsilon$  and  $\beta$  combine to give the modification of the beam dimension. The change of  $\alpha, \beta, \gamma$  also modifies the beam waist.

Exercise : Show that in a drift space the value  $\beta^*$  of the  $\beta$ -function at the waist and the distance  $l^*$  of the waist from a position where one knows the  $\alpha, \beta, \gamma$  parameters are given by :

$$\beta^* = \frac{\beta}{1 + \alpha^2} \text{ and } l^* = \frac{\beta\alpha}{1 + \alpha^2}.$$

- iv) The  $\alpha, \beta, \gamma$  parameters are not independent. From the equations 28 and 50 one obtains :

$$\beta\gamma = 1 + \alpha^2. \quad (51)$$

Imposing that relation when defining  $\alpha, \beta, \gamma$ , is equivalent to impose that the area of the envelope-ellipse is  $\pi\epsilon$ . The numerical factors in the definition of  $\alpha, \beta, \gamma$  and of  $\epsilon$  must be chosen in a fixed ratio to satisfy these conditions (47 and 51).

- v) In the case of a linear lattice, using the definition of  $\beta$ , one calculates its derivative:

$$\frac{d\beta}{ds} = \frac{4}{\epsilon} \frac{d}{ds} \overline{x^2} = \frac{8}{\epsilon} \overline{xx'},$$

i.e.

$$\alpha = -\frac{1}{2} \frac{d\beta}{ds}. \quad (52)$$

Similarly one obtains the derivative of  $\alpha$  and  $\gamma$  :

$$\begin{aligned} \frac{d\alpha}{ds} &= -\frac{4}{\epsilon} \frac{d}{ds} \overline{xx'} = -\frac{4}{\epsilon} (\overline{x'^2} - k\overline{x^2}) = k\beta - \gamma \\ \frac{d\gamma}{ds} &= \frac{4}{\epsilon} \frac{d}{ds} \overline{x'^2} = -\frac{8}{\epsilon} k\overline{xx'} = 2k\alpha. \end{aligned} \quad (53)$$

where  $k$  is the focusing strength along the lattice.

Now, in the traditional way, the  $\beta$ -function (and  $\alpha, \gamma$ ) are first introduced to describe the betatron motion of a particle in a linear and periodic lattice. Afterwards it appears that the  $\beta$ -function also represents the shape of the beam-envelope. In the statistical way, presented here, it is the reverse. The  $\beta$ -function is just defined to parametrize the beam-envelope. Therefore, one must show that it also describes the

betatron motion of a single particle. Ultimately both definitions of the  $\beta$ -function are equivalent. That last property of the statistical definition relies on the following property : "In a linear lattice, the envelope-ellipse at a position  $s_1$  is mapped into the envelope-ellipse of the new particle distribution at another position  $s_2$ . Moreover, any ellipse homothetic to the envelope-ellipse is mapped into an ellipse homothetic to the new envelope-ellipse".

That property can be proved in an elementary and algebraic way. At a first position  $s_1$  one writes the equation (48) of the envelope-ellipse, using a matrix form :

$$(x_1 x'_1) \overline{\begin{pmatrix} x'_1 \\ -x_1 \end{pmatrix}} (x'_1 - x_1) \begin{pmatrix} x_1 \\ x'_1 \end{pmatrix} = \frac{\epsilon^2}{4}. \quad (54)$$

At another position  $s_2$  the two column-vectors are mapped by a linear transformation  $\mathbf{R}$  :

$$\begin{pmatrix} x_2 \\ x'_2 \end{pmatrix} = \mathbf{R} \begin{pmatrix} x_1 \\ x'_1 \end{pmatrix}, \text{ and } \begin{pmatrix} x'_2 \\ -x_2 \end{pmatrix} = \mathbf{R}^{-1T} \begin{pmatrix} x'_1 \\ -x_1 \end{pmatrix}.$$

Inserting the unit products  $\mathbf{R}^T \mathbf{R}^{-1T}$  and  $\mathbf{R}^{-1} \mathbf{R}$ , the equation (54) of the envelope-ellipse at the position  $s_1$  can also be written :

$$(x_1 x'_1) \mathbf{R}^T \mathbf{R}^{-1T} \overline{\begin{pmatrix} x'_1 \\ -x_1 \end{pmatrix}} (x'_1 - x_1) \mathbf{R}^{-1} \mathbf{R} \begin{pmatrix} x_1 \\ x'_1 \end{pmatrix} = \frac{\epsilon^2}{4},$$

that becomes, by action of the operator  $\mathbf{R}$  on the adjacent column (or row) vectors, identical to the equation of the envelope-ellipse at the position  $s_2$ . It shows that the envelope-ellipse at the position  $s_1$  is mapped into the envelope-ellipse at the position  $s_2$ .

By changing the constant in the RHS of (49) and (54), one obtains the equation of any ellipse homothetic to the envelope-ellipse. Therefore, the preceding algorithm applies as well, proving the second statement of the property.

Now, a point representing a single particle belongs to an unique ellipse homothetic to the envelope-ellipse. That point is mapped into a point of the mapped ellipse, that is still homothetic to the envelope-ellipse, according to the latter property. One can describe the motion of the particle using a parametric representation of that ellipse. If  $J$  is the constant of the homothetic ellipse replacing  $\epsilon$  in the RHS of (48), the parametric representation is given by :

$$x(s) = \sqrt{\beta(s)J} \cos\phi(s) \quad (55)$$

$$x'(s) = -\sqrt{\frac{J}{\beta(s)}} \{ \sin\phi(s) + \alpha(s)\cos\phi(s) \}. \quad (56)$$

The angle  $\phi$  is identified to the betatron phase. Its variation is obtained by identifying the equation (56) to the derivative of (55) :

$$\frac{d\phi}{ds} = \frac{1}{\beta}. \quad (57)$$

In a transport line, the  $\beta$ -function is known when the particle distribution in phase space is known at the line entry. The  $\beta$ -function allows to describe that distribution and how it evolves along the line. It depends on the two initial values of  $\alpha$  and  $\beta$ <sup>[13,14]</sup>. In a periodic lattice, one can select a beam distribution that has the periodicity of the lattice. The beam is matched to the lattice. The  $\beta$ -function is the unique  $\beta$ -function that has this periodicity.

## REFERENCES

- [1] M. Weiss, CERN Accelerator School Proceedings (Editor : S. Turner)  
CERN 87-10 p 162-163 (1987)
- [2] P.M. Lapostolle, IEEE Trans. Nucl. Sci. NS-18, No 3, 1101 (1971)
- [3] A.H. Sørensen, CERN Accelerator School Proceedings (Editor : S. Turner)  
CERN 89-05 p 18-36 (1989)
- [4] G. H. Rees, CERN Accelerator School Proceedings (Editors : P. Bryant, S. Turner)  
CERN 85-19; vol. II p 346 (1984)
- [5] E. N. Wilson, CERN Accelerator School Proceedings (Editor : S. Turner)  
CERN 87-03 ; vol. I p 41 (1987)
- [6] J. Le Duff, CERN Accelerator School Proceedings (Editors : P. Bryant, S. Turner)  
CERN 85-19 ; vol. I p 125 (1984)
- [7] G. Dôme, CERN Accelerator School Proceedings (Editor : S. Turner)  
CERN 87-03 ; vol. I p 110 (1987)
- [8] E. Ciapala, CERN Accelerator School Proceedings (Editors : P. Bryant, S. Turner)  
CERN 85-19 ; vol. I p 195 (1984)
- [9] L. R. Evans & J. Gareyte, CERN Accelerator School Proceedings  
(Editor : S. Turner) CERN 87-03 ; vol. I p 159 (1987)
- [10] I. Hofmann, CERN Accelerator School Proceedings (Editor : S. Turner)  
CERN 87-03 ; vol. I p 337-338 (1987)
- [11] M. Sands, SLAC Report, SLAC/AP-47 (1985).
- [12] S.P. Møller, CERN Accelerator School Proceedings (Editor : S. Turner)  
CERN 89-05 p 1-17 (1989).
- [13] K. Steffen, CERN Accelerator School Proceedings (Editor : S. Turner) CERN  
85-19 p 43 (1985).
- [14] P. Bryant, CERN Accelerator School Proceedings (Editor : S. Turner) CERN  
85-19 p 361 (1985).



# LUMINOSITY MEASUREMENTS AND CALCULATIONS

*K. Potter*  
CERN, Geneva, Switzerland

## Abstract

The luminosity of a single-ring electron-positron collider is defined and an expression in terms of the beam intensity and geometry obtained. The equivalent expression for the case of coasting beams with a finite crossing angle is also derived. The use of the root mean square as a measure of particle density distributions is discussed. Methods of measuring the luminosity are described and indications of the important parameters for the accuracy of the measurements given.

## 1. INTRODUCTION

After the beam energy the most important parameter at a colliding beam facility, as far as the high energy physics user is concerned, is the counting rate. This is usually expressed by the term luminosity, which I shall define in a moment, but which as one can guess indicates the brilliance of the source. I shall first of all discuss the expressions which allow the collider builder to know what the luminosity will be in terms of his machine parameters. It is of course beyond the scope of this note to discuss the variation of these parameters with a view to increasing the luminosity.

Once the machine is built and operating the high energy physicist needs to know the machine luminosity in order to be able to normalise his measurements and he usually requires a more accurate value than can be obtained directly from machine parameters. I shall therefore go on to discuss methods of measuring luminosity from the experimenters' point of view.

## 2. DEFINITION OF LUMINOSITY

At a colliding beam facility the total interaction rate depends on the geometry of the beams, their density and energy, but the last is usually fixed by other requirements.

Consider first the interaction of a beam with a target of length  $l$  and particle number density  $n_2$ , as sketched in Fig. 1. Then the number of interactions ( $R$ ) per beam particle is proportional to  $n_2 \times l$  the total number of particles it can collide with and the constant of proportionality is defined as the cross section  $q$  for the type of interaction concerned.

$$R = q n_2 l \quad (1)$$

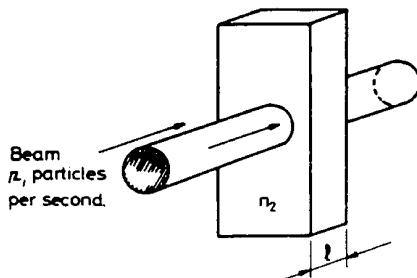


Fig. 1 Schematic of a particle beam of  $n_1$  particles per second incident

on a stationary target with  $n_2$  particles per unit volume

where  $q$  has the dimension  $\text{cm}^2$ . The transverse dimensions of the beam and target do not enter as the target is assumed to be wider than the beam. If the beam consists of  $n_1$  particles per second the rate of interactions

$$\frac{dR}{dt} = q n_1 n_2 l. \quad (2)$$

All the characteristics of the incident beam and target can be combined into a single term defined as the luminosity  $L$  by writing

$$\frac{dR}{dt} = q L \quad (3)$$

where  $L = n_1 n_2 l$  and has the dimensions  $\text{cm}^{-2} \text{s}^{-1}$ . Hence luminosity is simply the interaction rate per unit crosssection

### 3. LUMINOSITY OF A SINGLE-RING COLLIDER

In a colliding beam machine the expression for  $L$  is more complicated because the target is moving and we cannot always assume that the target is wider than the beam.

In a single-ring collider ( $e^+ e^-$ ) the two beams circulate in opposite directions. Suppose we have  $N$  particles per beam and the beams have equal r.m.s. radii of  $\sigma_x$  (horizontal) and  $\sigma_z$  (vertical). The cross-sectional area of the beam is then  $2\pi\sigma_x\sigma_z$  and the number of positrons which one electron encounters in one turn of the machine, assuminq the beams follow identical paths, is

$$\frac{q}{2\pi\sigma_x\sigma_z}$$

where  $q$  is an effective cross section of the electron.

The total rate  $\frac{dR_T}{dt}$  if the revolution frequency is  $f$  is

$$\frac{dR_T}{dt} = \frac{NfqN}{2\pi\sigma_x\sigma_z} \quad (4)$$

We have not said anything yet about the azimuthal distribution of the particles around the machine and for continuous or DC beams these interactions would be spread around the whole machine circumference. Electrons must be bunched for other reasons and if we have  $K$  bunches in each beam there are  $2K$  crossing points around the machine hence the interaction rate per crossing  $dR/dt$  is given by

$$\frac{dR}{dt} = \frac{N^2 f q}{4\pi k \sigma_x \sigma_z} \quad (5)$$

From our previous definition of luminosity in (3)

$$L = \frac{N^2 f}{4\pi k \sigma_x \sigma_z} \quad (6)$$

or since it is more usual to measure beam currents  $I$  where  $I = N \cdot e \cdot f$

$$L = \frac{I^2}{4\pi kf\sigma_x\sigma_z} \quad (7)$$

$\sigma_x$  and  $\sigma_z$  are the r.m.s. dimensions of the two beams at the crossing points and we have assumed that they are constant over the effective crossing region which is of course a function of the bunch length. We have also assumed identical positron and electron beams which is normally the case. If one beam has a much larger cross section than the other, has been blown up by the beam-beam interaction for example, then the  $\sigma_x\sigma_z$  term must be replaced by an effective area  $A_{\text{eff}}$  which will be approximately the area of the larger beam. In the case of a proton-antiproton collider there is no reason to expect equal beams and the same remarks apply.

As an example of a more complicated formulation of these geometric factors, that given [1] for the electron proton ring HERA is perhaps instructive

$$L = \frac{f \cdot N_p \cdot N_e}{k \cdot 2\pi(\sigma_{xp,\text{eff}}^2 + \sigma_{xe}^2)^{1/2} \cdot (\sigma_{zp}^2 + \sigma_{ze}^2)^{1/2}} \quad (8)$$

where  $\sigma_{xp,\text{eff}} = (\sigma_{xp}^2 + (\sigma_{sp} \cdot \phi)^2)^{1/2}$ ,  $\sigma_{sp}$  is the proton bunch length with  $\phi$  the half crossing angle ( $\pm 5$  mrad).  $\sigma_{xp}$  and  $\sigma_{xe}$  are the r.m.s. widths of the proton and electron beam respectively and  $\sigma_{zp}$  and  $\sigma_{ze}$  are the r.m.s. heights.

#### 4. THE RMS AS A MEASURE OF BEAM HEIGHT

In all cases it has been assumed that the r.m.s. value of the beam density distribution is a good measure of beam size. It is interesting to consider to what extent this is true for different distributions or are we implying always a Gaussian distribution? To examine this further consider the simple case of two coasting beams interacting at an angle in one plane and with identical distributions  $\rho(y)$  in the other plane. In the plane of the crossing there is no vertical dispersion and all particles of one beam intersect all particles of the other beam, hence the luminosity depends only on the distribution  $\rho(y)$ . For convenience we can consider  $\rho(y)$  as normalised to unity so that:

$$\int_{-\infty}^{\infty} \rho(y) dy = 1. \quad (9)$$

The mean square width of this distribution is then

$$\langle y^2 \rangle = \int_{-\infty}^{\infty} y^2 \rho(y) dy \quad (10)$$

For given circulating currents (fixed numbers of particles) the interaction rate in this simple case is proportional to:

$$L = \int_{-\infty}^{\infty} \rho^2(y) dy \quad (11)$$

which we notice means that  $L = \langle \rho \rangle$  since the general definition of an average is

$$\langle f \rangle = \int_{-\infty}^{\infty} f(y) \rho(y) dy \quad (12)$$

In other words  $L$  is the density of the target beam, as seen on average by the bombarding beam.

For a given distribution, we have seen (Eq. (6)) that  $L$  is inversely proportional to the width of the distribution parametrised by the r.m.s. width  $\langle y^2 \rangle^{1/2}$ .

Therefore  $L \langle y^2 \rangle^{1/2}$  must be a constant. Suppose  $\rho$  is a Gaussian distribution

$$\rho = \frac{1}{\sigma\sqrt{2\pi}} \exp^{-y^2/2\sigma^2} \quad (13)$$

then using (10)

$$\langle y^2 \rangle = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} y^2 \exp^{-y^2/2\sigma^2} dy = \sigma^2 \quad (14)$$

and

$$L = \frac{1}{2\pi\sigma^2} \int_{-\infty}^{\infty} \exp^{-y^2/\sigma^2} dy = \frac{1}{2\sigma\sqrt{\pi}} \quad (15)$$

and

$$L \langle y^2 \rangle^{1/2} = \frac{1}{2\sqrt{\pi}} = 0.28209$$

The interesting observation made by H. Hereward [2] is that the same constant is approximately valid for many different distributions. For example consider a rectangular distribution defined by

$$\rho = \frac{1}{2y_0} \text{ for } -y_0 \leq y \leq y_0 \quad (16)$$

Hence

$$\langle y^2 \rangle = \frac{1}{2y_0} \int_{-y_0}^{+y_0} y^2 dy = \frac{1}{3} y_0^2 \quad (17)$$

and

$$L = \frac{1}{4y_0^2} \int_{-y_0}^{+y_0} dy = \frac{1}{2y_0}. \quad (18)$$

Therefore

$$L \langle y^2 \rangle^{1/2} = \frac{1}{2} \sqrt{\frac{1}{3}} = \frac{\sqrt{3}}{6} = 0.28868$$

Similarly for the distributions of Table 1.

The range of the results given in the last column of Table 1 is only  $\pm 3.7\%$ . In fact Hereward [2] was able to show that  $L \langle y^2 \rangle^{1/2}$  is never less than 0.2683 (a parabolic distribution) for any distribution  $\rho \geq 0$ . Unfortunately he also showed that there is no equivalent limit to the maximum value and the distribution illustrated in Fig. 2 can have an arbitrarily large  $L \langle y^2 \rangle^{1/2}$ .

**Table 1**

$L \langle y^2 \rangle^{1/2}$  for some typical distributions

Distribution within $-y_0 \leq y \leq y_0$	$L \langle y^2 \rangle^{1/2}$
Parabola $\rho = \frac{3}{4y_0} \left( 1 - \frac{y^2}{y_0^2} \right)$	$\frac{3\sqrt{5}}{25} = 0.2683$
Cosine $\rho = \frac{\pi}{4y_0} \cos \frac{\pi y}{2y_0}$	$\frac{\pi^2}{16} \sqrt{1 - \frac{8}{\pi^2}} = 0.2685$
Triangle $\rho = \frac{1}{y_0} \left( 1 - \left  \frac{y}{y_0} \right  \right)$	$\frac{2}{3} \frac{1}{\sqrt{6}} = 0.2722$
Truncated Gaussian $\rho = \frac{1}{\sigma\sqrt{2\pi} \alpha\left(\frac{y_0}{\sqrt{\sigma}}\right)} \exp^{-y^2/2\sigma^2}$ where $\alpha(x)$ is the "normal probability integral" $[\alpha(x) = \frac{1}{\sqrt{2\pi}} \int_{-x}^x \exp^{-t^2/2} dt]$	Limit $\frac{y_0}{\sqrt{\sigma}} \rightarrow 0, \quad \frac{\sqrt{3}}{6} = 0.2887$ Limit $\frac{y_0}{\sqrt{\sigma}} \rightarrow \infty, \quad \frac{1}{2\sqrt{\pi}} = 0.2809$ Minimum $\frac{y_0}{\sqrt{\sigma}} (1.63), \quad = 0.2694$

It can be argued that particle beams are much more likely to be something like a truncated Gaussian or one of the other distributions of Table 1 than that of Fig. 2. In particular since gas scattering or other random kick processes increase the width of a narrow beam making it more and more Gaussian.

## 5. LUMINOSITY WITH COASTING BEAMS AND FINITE CROSSING ANGLE

Colliding beam machines with two separate rings and a crossing angle can operate with coasting, that is to say unbunched beams. The interaction region length is then defined by the beam dimension in the plane of the crossing angle as illustrated in Fig. 3 for the ISR. We can again show that the luminosity is inversely proportional to the beam height.

In order to derive the luminosity formula assume two ribbon beams of height  $h$  and width  $w$  crossing in the horizontal plane with angle  $\alpha$  as in Fig. 3. We first note that since the number of particles encountered must be the same for all observers we can choose any convenient system such as the rest frame of beam 2. From Eq. (1) we know that we have to determine the effective particle density and the length of traversal of the test particle  $Q$  of beam 1.

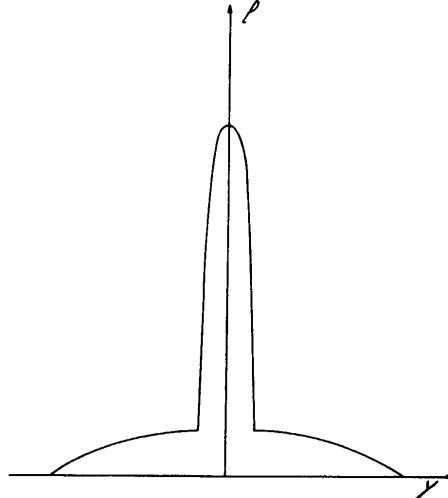


Fig. 2 A particle distribution that can have  $L \langle y^2 \rangle^{1/2}$  arbitrarily large. A fraction of the particles are spread out over a constant width and contribute a constant amount of  $\langle y^2 \rangle$ , while the remaining fraction is in a peak that contributes increasingly to  $L$  as it is made narrower,

$$L = \int_{-\infty}^{\infty} \rho^2(y) dy .$$

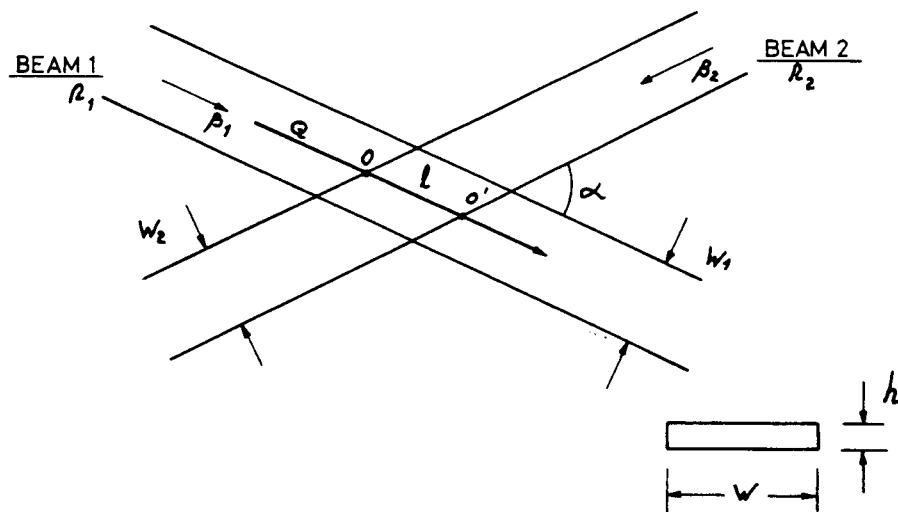


Fig. 3 Schematic of coasting beams of width  $W$  and height  $h$  colliding with a crossing angle  $\alpha$  in the horizontal plane

The Lorentz transformation needed to bring beam 2 to rest when applied to particle  $Q$  of beam 1 changes the angle of traversal  $\alpha$  to  $\alpha'$  given by

$$\tan \alpha' = \frac{\sin \alpha}{\gamma_2 \left( \cos \alpha + \frac{\beta_2}{\beta_1} \right)} \quad (19)$$

where  $\beta_1$  and  $\beta_2$  are the usual ratios of the velocities of beam 1 and beam 2 respectively to that of light ( $\beta = v/c$ ) and  $\gamma = (1 - \beta^2)^{-1/2}$ .

The length of traversal  $l'$  is then  $W_2 / \sin \alpha'$ .

Since for equal energy beams  $\beta_1 = \beta_2 = \beta$  and  $\gamma_1 = \gamma_2 = \gamma$ , Eq. (19) becomes

$$\tan \alpha' = \frac{\sin \alpha}{\gamma(\cos \alpha + 1)} = \frac{\tan \alpha/2}{\gamma} \quad (20)$$

and by simple trigonometry

$$\sin \alpha' = \frac{\tan \alpha/2}{\gamma} \sqrt{1 + \frac{\tan^2 \alpha/2}{\gamma^2}}$$

For high energy beams  $\beta \sim 1$  and  $\gamma$  is large so that the second term can be neglected, giving

$$\sin \alpha' = \frac{\tan \alpha/2}{\gamma}$$

and

$$l' = \frac{W_2 \gamma}{\tan \alpha/2} \quad (20)$$

The unit volume of beam 2 is also reduced by the Lorentz contraction so that the particle density ( $n_2'$ ) seen by  $Q$  is given by

$$n_2' = \frac{n_2}{\gamma} \quad (21)$$

Using (20) and (21) in Eq. (1) gives directly the number of interactions per traversal as

$$q \cdot \frac{n_2 w_2}{\tan \alpha/2} \quad (22)$$

The number of particles per second in beam 1 is

$$n_1 v_1 w_1 h$$

and the total interaction rate is therefore

$$\frac{dN}{dt} = q \cdot \frac{n_1 n_2 v_1 w_1 w_2 h}{\tan \alpha/2} \quad (23)$$

in terms of current

$$n = \frac{I}{ewhc} \text{ again assuming } v = c$$

$$\frac{dN}{dt} = \frac{\sigma}{ce^2} \frac{I_1 I_2}{h \tan \alpha/2} \quad (24)$$

Since we have two rings the beams can have different heights and profiles and are not necessarily well aligned, so that  $h$  must be replaced by  $h_{\text{eff}}$  where

$$h_{\text{eff}} = \frac{\int_{-\infty}^{\infty} \rho_1 dz \int_{-\infty}^{\infty} \rho_2 dz}{\int_{-\infty}^{\infty} \rho_1 \rho_2 dz} \quad (25)$$

Using Eqs. (3) and (24) the luminosity  $L$  becomes

$$L = \frac{I_1 I_2}{ce^2 h_{\text{eff}} \tan \alpha/2} \quad (26)$$

and once again the problem of determining the luminosity is essentially that of determining the beam geometry term  $h_{\text{eff}}$ .

## 6. MEASUREMENT OF THE LUMINOSITY

In principle Eq. (7) implies that at least in the case of a single-ring collider the luminosity is known provided the transverse beam dimensions can be measured. However, it is clear that if the beam profiles are measured elsewhere in the machine they must be transformed to the interaction point using a knowledge of  $\beta_y$  at the point of measurement and at the beam crossing point.

In practice, while machine designers make use of such expressions, at an operating collider the experimenters require a more precise knowledge of the luminosity for normalization purposes and must find other means. At an electron machine the standard technique is to set up a monitor consisting of two small-angle electron telescopes to observe elastic (Bhabha) scattering as in Fig. 4. This process has a well-known, exactly calculable, cross section,  $\sigma_B$  so that from a measurement of the counting rate the luminosity can be determined using Eq. (3). The  $\sigma$  must of course be calculated for the acceptance of the detector

$$\sigma = \int_{\theta_{\min}}^{\theta_{\max}} \frac{d\sigma_B}{d\Omega} d\Omega \quad (27)$$

with

$$\frac{d\sigma_B}{d\Omega} = \frac{\alpha^2}{8E^2} \frac{(2 - \sin^2 \theta)(4 - \sin^2 \theta)}{\sin^4 \theta} \quad (28)$$

from quantum electrodynamics, where  $\theta_{\min}$  and  $\theta_{\max}$  are the minimum and maximum scattering angles accepted by the monitor,  $E$  is the beam energy and  $\alpha$  is the fine structure constant ( $e^2/hc$ ). If care is taken to choose a counter configuration which minimises the effects

of interaction region displacements and other geometrical effects the luminosity can be determined to a few per cent [3]. But, since counting rates are low, less sophisticated monitors with larger counters are often used to give rapid values of the luminosity to  $\sim 10\%$  accuracy. These monitors are then very useful diagnostic devices of the machine, giving information on beam sizes via Eq. (7) for example.

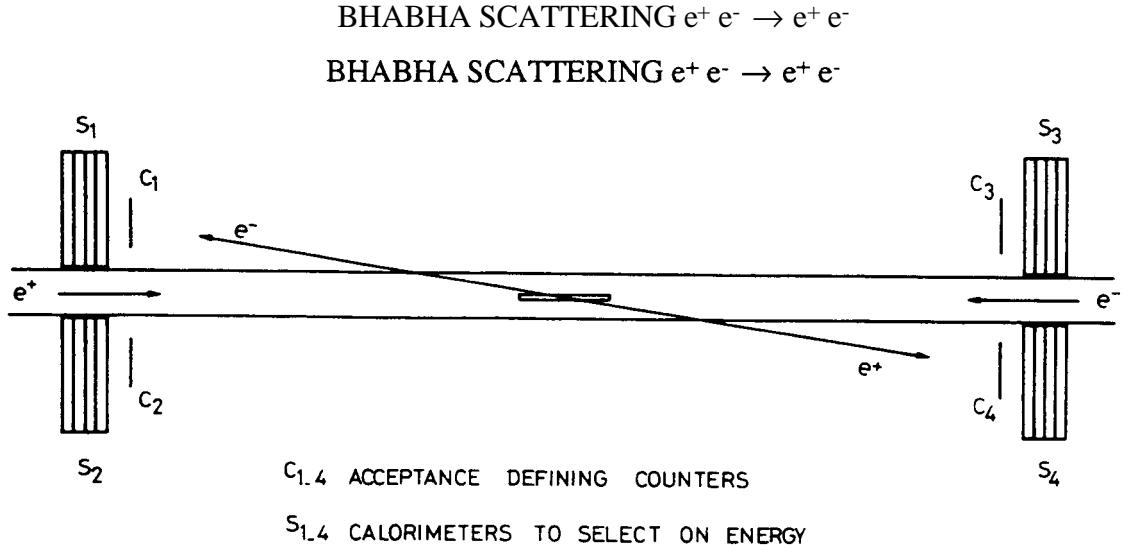


Fig. 4 Layout of a typical luminosity monitor to detect Bhabha scattering at electron-positron storage rings

## 7. THE VAN DER MEER METHOD OF LUMINOSITY MEASUREMENT

For two-ring colliders Eq. (25) allows a calculation of  $h_{\text{eff}}$  from a knowledge of the vertical beam profiles but for the evaluation of the numerator the vertical distance between the centres of the beam distributions at the crossing must be known. In the ISR it was out of the question to use the standard bunched beam pick-up system to give this, mainly because of the dependence of the closed orbit on intensity as a result of space charge forces. The only possibility was to steer the beams vertically while observing a suitable monitor counting rate in order to maximise the luminosity and obtain the  $h = 0$  point.

This idea of vertical beam steering enabled S. van der Meer [4] to point out a much cleverer way of measuring the luminosity.

If one measures the counting rate in a luminosity monitor, of a similar layout to that of Fig. 4, as a function of relative vertical separation ( $h$ ) of the two beams, a curve similar to that of Fig. 5 will result with a maximum at zero separation. Van der Meer showed that the area under this curve divided by the value at  $h = 0$  is  $h_{\text{eff}}$ .

At separation  $h$  the counting rate is

$$A \cdot \int \rho_1(z) \cdot \rho_2(z - h) dz \quad (29)$$

where  $A$  is an unknown constant which includes the interaction cross section and the acceptance of the monitor but which can be assumed constant at fixed energy.

Then the area under the counting rate curve is

$$\int [A \int \rho_1(z) \cdot \rho_2(z - h) dz] dh = A \int [\rho_1(z) \int \rho_2(z - h) dh] dz \quad (30)$$

and the rate at  $h = 0$  is

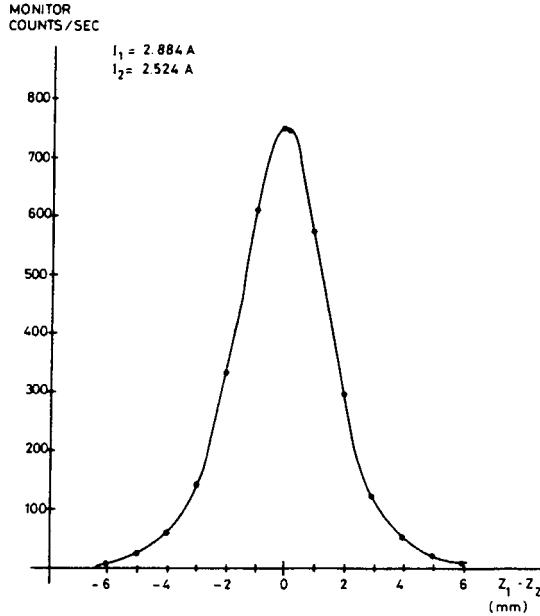


Fig. 5 The counting rate in a luminosity monitor as a function of the vertical separation ( $z_1 - z_2$ ) between the two coasting beams.

$$A \cdot \int \rho_1(z) \cdot \rho_2(z) dz \quad (31)$$

If the integrals are taken over the entire non-zero region then

$$\int \rho_2(z - h) dh = \int \rho_2(z) dz$$

and therefore Eq. (30) divided by (31) is

$$\frac{\int [\rho_1(z) \int \rho_2(z) dz] dz}{\int \rho_1(z) \cdot \rho_2(z) dz} = \frac{\int \rho_1(z) dz \int \rho_2(z) dz}{\int \rho_1(z) \cdot \rho_2(z) dz} = h_{\text{eff}} \quad (32)$$

With  $h_{\text{eff}}$  determined the luminosity can be calculated using Eq. (26). The currents  $I_1$  and  $I_2$  can be very accurately measured using a DC current transformer [5] and the crossing angle of the beams is known to high precision.

Since from the definition of luminosity the counting rate ( $R_M$ ) in a monitor is given by

$$R_M = \sigma_M \cdot L$$

the monitor cross section  $\sigma_M$  has in effect been measured by the above procedure and to the extent that the monitor rate is unaffected by backgrounds, geometry of the crossing region and its efficiency does not change, this monitor rate can be used as a direct measure of luminosity.

At the ISR this technique worked extremely well and with occasional calibrations of their monitors the experimenters always knew the luminosity to within a few per cent. For particular experiments such as the measurement of the total p-p and p- $\bar{p}$  cross section special care was taken and an error of less than 1% was achieved. In particular this required a calibration of the beam displacement ( $h$ ) used in the luminosity measurement.

## 8. VERTICAL BEAM DISPLACEMENTS AT THE ISR

The most convenient way of creating a local closed orbit bump is to place two dipole magnets a quarter of a wavelength before and after the crossing point. In practice, however, the phase advances cannot be exact because of lack of space or simply because the tune of the machine is not fixed. At the ISR this problem was solved by using the horizontal field steering magnets of the adjacent intersecting regions to make the necessary corrections. The orbit distortion was then as illustrated by Fig. 6. The problem was to ensure that the one millimetre nominal displacement at the intersection was correct to a few parts per thousand and was reproducible to a similar level. This is a much better precision than is obtainable with the usual programs such as AGS [6] or MAD [7] and in addition great care with power supply setting and magnet hysteresis was required.

The solution was to calibrate the beam displacements using a scraper driven by a precision screw [8]. At the ISR special scrapers [9] were used for obtaining vertical beam profiles but as in this case only the precise centre of the distribution was required the technique was simplified so that it amounted to giving the beam a sharp edge with say the upper scraper and then finding that edge with the lower scraper. By repeating this procedure many times a very precise determination of the beam centre was possible. The location of the beam centre to  $\pm 3 \mu\text{m}$  was more than adequate to establish the linearity of the displacements, the success of a hysteresis correction routine, and to provide an absolute calibration to within  $\pm 4 \%$ . Dependence of the bump amplitude on all sorts of parameters such as machine tune, beam intensity, betatron coupling and even the horizontal closed orbit were all studied in order to make sure that the displacement scale measured with the scrapers using single pulses was the same as for the beams of a few amperes which were used to measure the total cross sections.

## 9. CONCLUDING REMARKS

The use of the van de Meer method to measure  $h_{\text{eff}}$  and calibrate luminosity monitors was a great success at the ISR; where, as explained above, the luminosity depended only on the beam dimensions in one plane. In the more usual case of bunched beams, where both horizontal and vertical beam dimensions influence the luminosity it is in principle still possible to use the van der Meer method by scanning in both planes. However, the problems alluded to above become even more important and for finite crossing angles monitor acceptance as a function of longitudinal position has to be considered very carefully as the collision region is displaced as the beams are scanned in the plane of the crossing. In practice a combination of several methods including comparing small angle scattering with the total cross section, through the optical theorem, are used to establish the luminosity at colliders. For the next generation of hadron colliders such as the LHC and SSC there will be additional problems arising from the very small angles of elastic scattering in the 10 TeV region. Indeed most secondaries also stay in the beam pipe, which will not only make the use of the optical theorem very difficult but will also make the choice of a good luminosity monitor less than obvious. At this time it seems doubtful if the true luminosity of these future colliders will ever be known with an absolute accuracy better than a few percent.

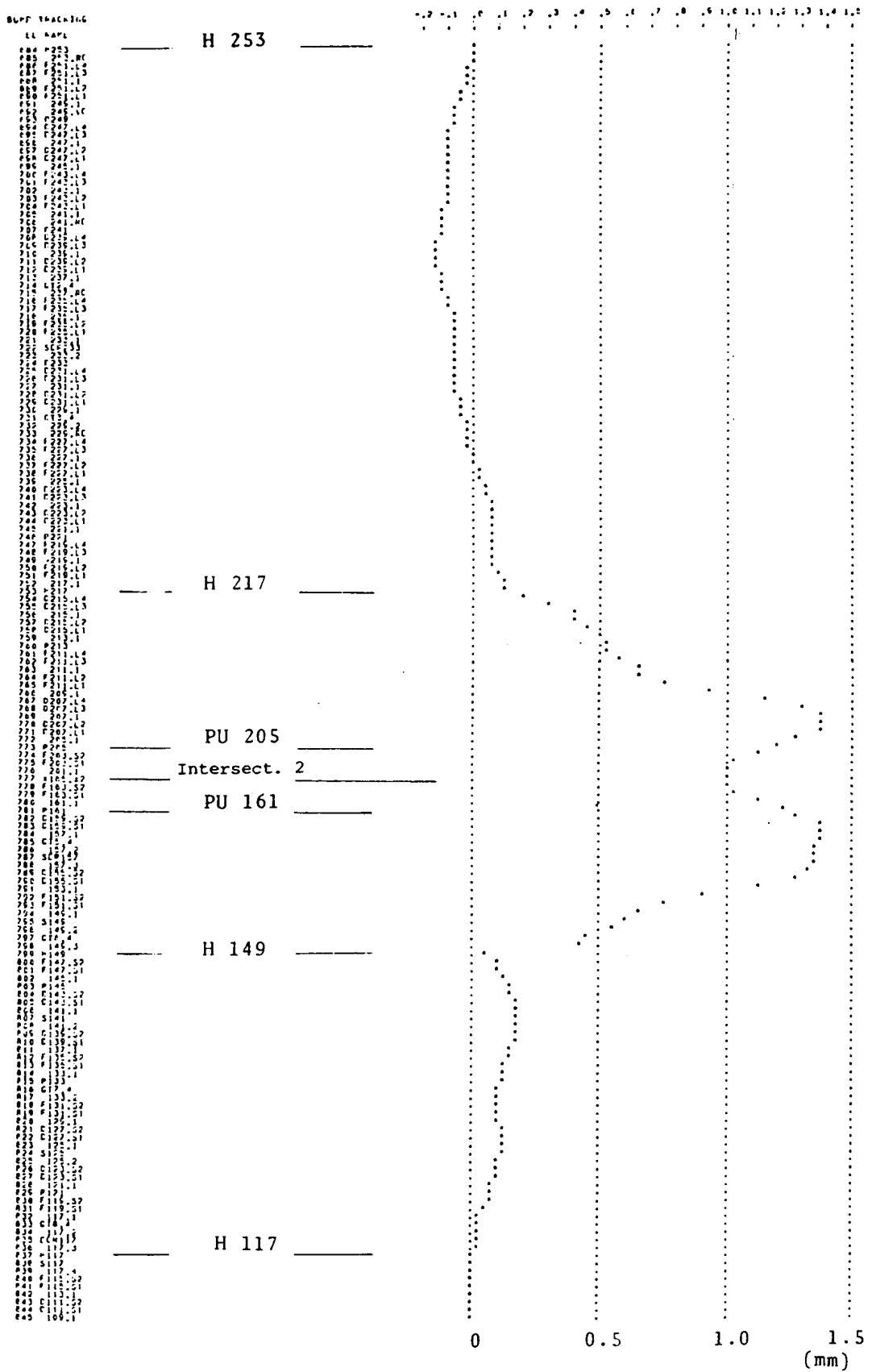


Fig. 6 An AGS tracking of a 1-mm closed orbit bump as

used for luminosity measurements at an ISR intersection.

## REFERENCES

- [1] Study on the Proton Electron Storage Ring Project HERA, ECFA 80/42 (17 March 1980), DESY HERA 80/01, page III-9.
- [2] H.G. Hereward, CERN Internal Report, How good is the r.m.s. as a measure of beam size? CERN/MPS/DL 69-15 (1969).
- [3] J.F. Crawford et al., A precision luminosity monitor for use at electron positron storage rings, Nucl. Instr. & Meth. 127 (1975) 173-182.
- [4] S. van der Meer, CERN Internal Report, Calibration of the effective beam height in the ISR, ISR PO/68-31, June 1968.
- [5] K. Unser, A toroidal DC beam current transformer with high resolution, Proceedings of the 1981 Particle Accelerator Conf. IEEE Trans. on Nucl. Sci. vol. NS-28 No 3 (1981) 2344.
- [6] E. Keil et al., AGS – The ISR computer program for synchrotron design, orbit analysis and insertion matching, CERN 75-13 (1975).
- [7] H. Grote and F.C. Iselin, The MAD program, CERN/SL/90-13 (AP) (Revision 3) (1992).
- [8] P. Bryant and K. Potter, Calibration of the beam displacements used in ISR luminosity measurements, CERN Internal Report, CERN-ISR-ES-BOM/82-15 (1982).
- [9] K. Potter and S. Turner, IEEE Transactions on Nuclear Science, Vol.NS-22, No. 3, p1589 (1975).

# TRANSVERSE BEAM DYNAMICS

*E. Wilson*  
CERN, Geneva, Switzerland

## Abstract

This contribution describes the transverse dynamics of particles in a synchrotron. It builds on other contributions to the General Accelerator School for definitions of transport matrices and lattice functions. After a discussion of the conservation laws which govern emittance, the effects of closed orbit distortion and other field errors are treated. A number of practical methods of measuring the transverse behaviour of particles are outlined.

## 1. INTRODUCTION

During the design phase of an accelerator project a considerable amount of calculation and discussion centres around the choice of the transverse focusing system. The lattice, formed by the pattern of bending and focusing magnets, has a strong influence on the aperture of these magnets which are usually the most expensive single system in the accelerator.

Figure 1 shows the lattice design for the SPS at CERN. The ring consists of a chain of 108 such cells. The principles of designing such a lattice have been treated elsewhere in this school by Rossbach and Schmüser [1] and the methods of modifying the regular pattern of cells to make matched insertions where space is needed for accelerating structures, extraction systems and for experiments where colliding beams interact, will be the subject of another talk [2].

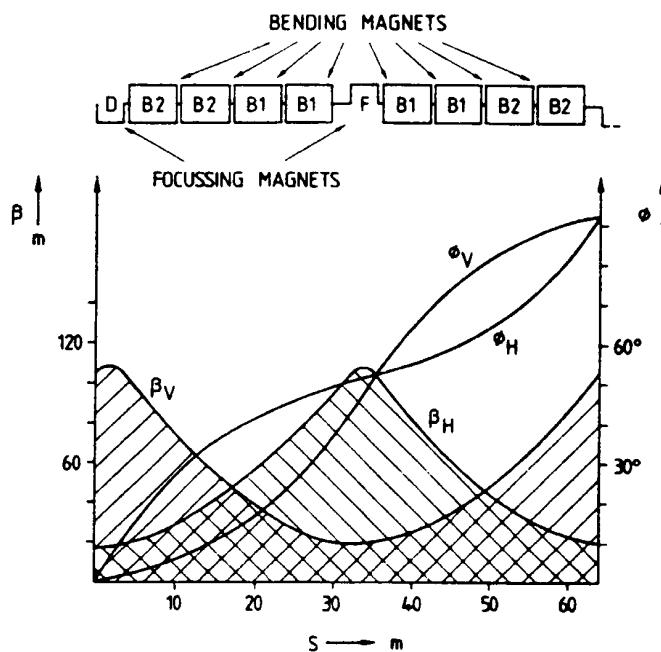


Fig. 1 One cell of the CERN 400 GeV Super Proton Synchrotron representing 1/108 of the circumference. The pattern of dipole and quadrupole magnets (F and D) is shown above. Beam particles make betatron oscillations within the shaded envelopes.

In this series of lectures I shall concentrate on those aspects of transverse dynamics which frustrate the designer or plague the person whose job it is to make the machine perform as well as theory predicts. I shall start with Liouville's theorem because this is an inescapable limitation to the beam size. Next I shall explain the distortion of the central orbit of the machine due to errors in the bending field. I shall then move on to explain how errors in quadrupole gradients may lead to instability of the betatron motion at certain  $Q$  values known as stopbands. Finally I shall explain why it is necessary to correct the variation of  $Q$  with momentum with sextupoles in order to avoid these stopbands.

The more advanced topics of the influence of fields which couple the two transverse phase planes together and fields whose non-linear nature can excite resonances which extract particles are given in two separate talks in this school [3,4]. In those talks we shall move away from the predictable linear behaviour to glimpse the fascinating jungle of effects which stretch our minds at the boundaries of accelerator theory and which will be fully developed in the Advanced Course of this school.

Since the effects I shall mention are annoying and best eliminated I shall try to indicate how measurements may be made on a circulating beam and its transverse behaviour corrected. Many of the effects I shall describe need only to be estimated rather than accurately predicted but it is very important to understand their mechanism. I shall therefore prefer to introduce concepts with approximate but physically revealing theory.

I make no apologies for building the theory from elementary definitions. These have to be restated to give those not fortunate enough to have eaten and slept in phase space a firm basic understanding. I hope that others who become impatient with the redefinition of basic quantities will be eventually gratified by the later stages in the exposition of the theory and the references I have given for further study.

## 2. LIOUVILLE'S THEOREM

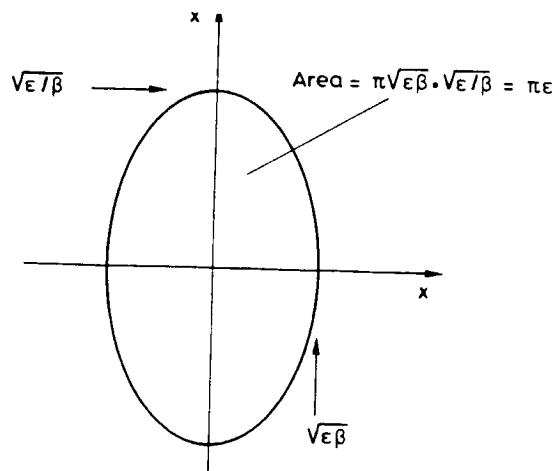


Fig. 2 Liouville's theorem applies to this ellipse

Particle dynamics obey a conservation law of phase space called Liouville's theorem. A beam of particles may be represented in a phase space diagram as a cloud of points within a closed contour, usually an ellipse (Fig. 2).

The area within the contour is proportional to the emittance of the beam. At constant energy we express this as the area  $\epsilon = \int y' dy$  in units of  $\pi \text{mm.milliradians}$ . We have shown the contour at a place where the  $\beta$  function is at a maximum or minimum and where the major and minor axes of an upright ellipse are  $\sqrt{\epsilon\beta}$  and  $\sqrt{\epsilon/\beta}$ .

This emittance is conserved, whatever magnetic focusing or bending operation we do on the beam. This is a consequence of the conservation of transverse momentum or, more strictly, a consequence of Liouville's theorem which states that:

*"In the vicinity of a particle, the particle density in phase space is constant if the particles move in an external magnetic field or in a general field in which the forces do not depend upon velocity."*

Clearly this definition rules out the application of Liouville's theorem to space charge problems or when particles emit synchrotron light. However, with these exceptions, it allows us to be confident that though the beam cross section may take on many shapes around the accelerator, its phase space area will not change (Fig. 3). Thus if the beam is at a narrow waist (a) its divergence will be large. In an F quadrupole (d) where the betatron function is maximum, its divergence will be small and a small angular deflection can greatly dilute phase space.

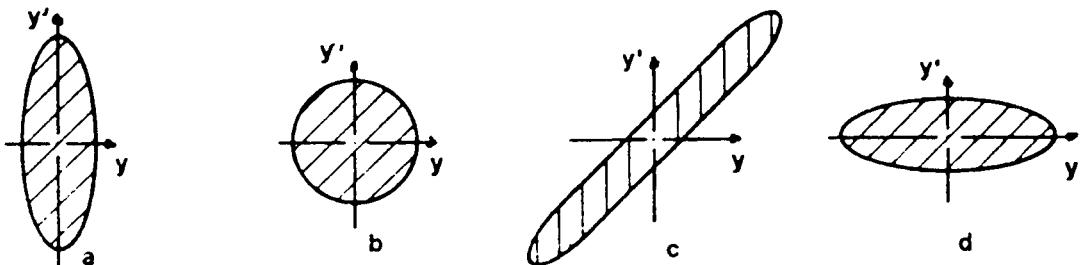


Fig. 3 The development of a constant emittance beam in phase space at a) a narrow waist, b) a broad waist or maximum in the beta function, c) place where the beam is diverging and d) at a broad maximum at the centre of an F lens.

In regions where the  $\beta$  function is not at a maximum or minimum the ellipse is tilted representing a beam that is either diverging or converging. In Fig. 4 we see the ellipse portrayed in terms of the Twiss parameters. The equation of the ellipse (often called the Courant and Snyder invariant) is:

$$\gamma(s)y^2 + 2\alpha(s)yy' + \beta(s)y'^2 = \epsilon .$$

and the invariance of this quantity as we move to different points in the ring is a consequence of Liouville's theorem.

All of this is true for a beam circulating at constant energy, but we may go further and apply Liouville's theorem to a proton beam during acceleration. To do so we must express emittance in the canonical phase space ( $q, p$ ), and we must first look carefully at the coordinates: displacement,  $y$ , and divergence,  $y'$ , we have been using so far. The canonical coordinates of relativistic mechanics are:

$$q = y$$

$$p = \frac{m\dot{y}}{\sqrt{1 - v^2/c^2}} = mc\gamma\beta_y$$

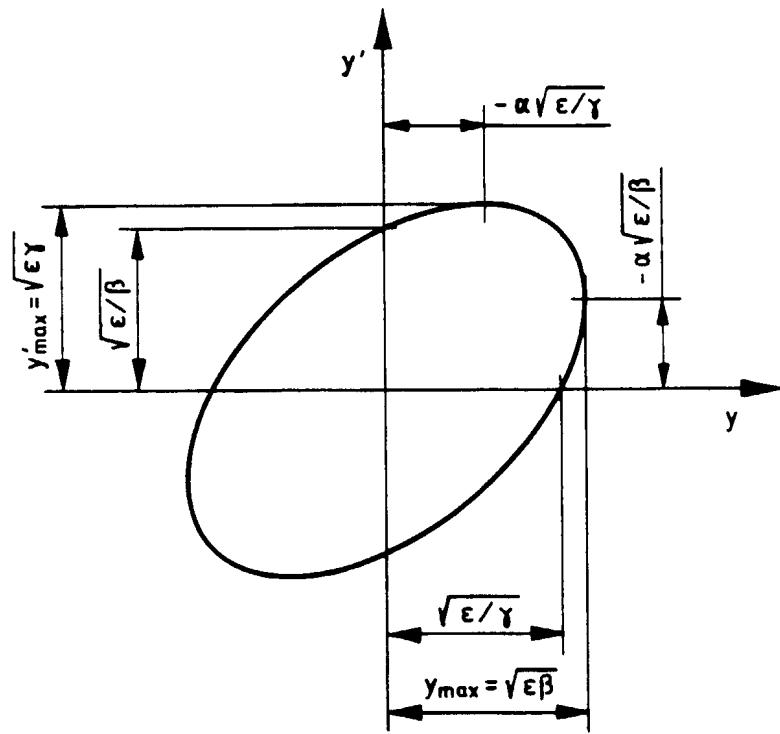


Fig. 4 The parameters of a phase-space ellipse containing an emittance  $\epsilon$  at a point in the lattice between quadrupoles

where we are obliged to redefine the meaning of  $\beta$  and  $\gamma$ :

$$\begin{aligned}\beta_y &= \dot{y}/c, \\ \gamma &= \left[1 - (\dot{x}^2 + \dot{z}^2 + \dot{s}^2)/c^2\right]^{-1/2}, \\ m &= \text{rest mass}, \\ c &= \text{velocity of light}.\end{aligned}$$

The divergence is therefore:

$$y' = \frac{dy}{ds} = \frac{\dot{y}}{\dot{s}} = \frac{\beta_y}{\beta},$$

where

$$\beta = v/c.$$

and the canonically conjugate  $p$  and  $q$  are therefore related to  $y$  and  $y'$ :

$$\begin{aligned}p &= mc\beta\gamma \cdot y' \\ q &= y\end{aligned}$$

In its general form Liouville's theorem states:

$$\int p \, dq = \text{constant} = mc\beta\gamma \int y' \, dy = mc\beta\gamma \cdot \pi\epsilon = p_0 c \cdot \pi\epsilon$$

where:

- $p_0$  is the momentum of the particle,
- $\varepsilon$ , the emittance, which is invariant at any given energy.

Accelerator physicists often quote an invariant emittance:

$$\varepsilon^* = \beta\gamma\varepsilon [\pi \cdot \text{mm.mrad}] .$$

As acceleration proceeds the normalized emittance is conserved;  $\beta\gamma$  is just proportional to the momentum or, above a few GeV, to the energy of the proton. Thus, we expect the emittance  $\varepsilon$  to shrink "adiabatically" during acceleration as  $1/p_0$  and the beam dimensions to shrink as  $1/p_0^{1/2}$  (Fig. 5).

As a rule, proton accelerators need their full aperture at injection and it is then that their design is most critical. It is for this reason too, that multistage accelerators such as the Linac-PSB-PS complex are used, since by inserting the PSB the energy of the Linac beam is increased, thus allowing a beam of larger normalized emittance  $\varepsilon^*$  and containing more protons, to be injected into the PS.

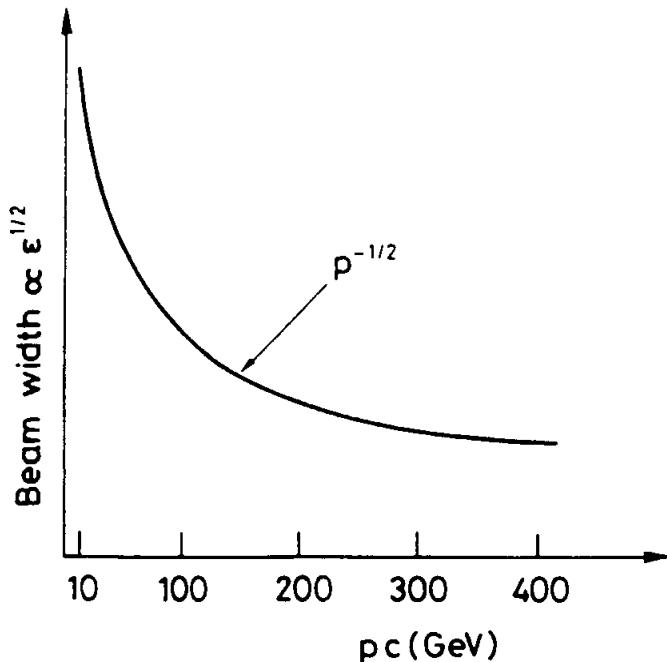


Fig. 5 Adiabatic shrinking

## 2.1 Chains of accelerators

In a chain of proton accelerators such as those at FNAL or CERN the invariant emittance, fixed by space charge at the beginning of the linac may be conserved to several hundred GeV.

Of course, in practice there may be mismatches between machines or non-linear fields which dilate the emittance. There are even techniques, stochastic and electron cooling, which condense the emittance and appear to defeat Liouville's theorem. In fact they merely rearrange and centralise the particles within the emittance contour rather than affecting the mean density.

Another, more mathematical, description of Liouville's theorem is to consider the transformation of a small element of phase space bounded by eight particles, including a ninth and excluding a tenth particle. Suppose the two diagrams of Fig. 6 are related by the transformation,  $M$ . Then, in this picture area is related by the Jacobian determinant:

$$J = |M| = \begin{vmatrix} \frac{\partial y}{\partial y_0} & \frac{\partial y}{\partial y'_0} \\ \frac{\partial y'}{\partial y_0} & \frac{\partial y'}{\partial y'_0} \end{vmatrix} = 1 .$$

Since the determinant is 1, area is conserved and Liouville follows. In other words Liouville's theorem implies that the determinant of the transport matrix is unity.

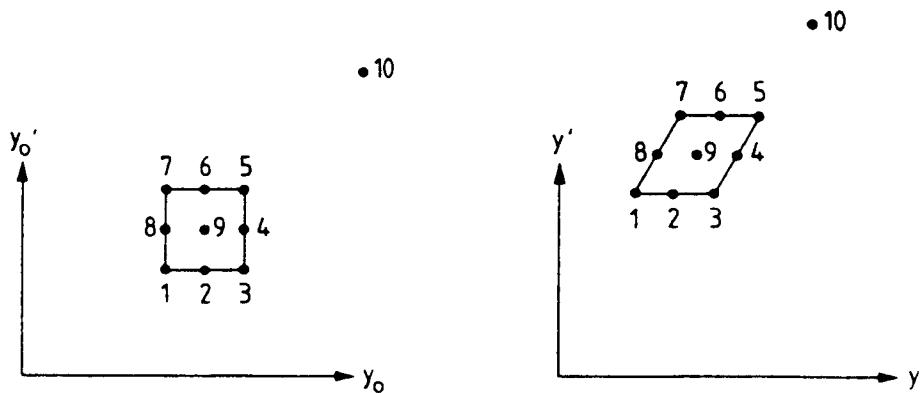


Fig. 6 The development of 10 points in phase space undergoing a Liouvillian transformation

## 2.2 Exceptions to Liouville's theorem

By now, anyone who comes from a laboratory that has an electron synchrotron will be beginning to experience a rising tide of protest. In fact, the invariance of normalised emittance and the shrinking of physical emittance with energy is quite the opposite of what happens in an electron machine.

As we shall read in later contributions on synchrotron radiation [5], electrons, being highly relativistic, emit quanta of radiation as they are accelerated. The size of the average quantum increases with  $E^3$  and the rate of emission as  $E$ . This quantized emission causes particles to jump around in momentum, changes which couple into transverse phase space. At the same time there is a steady tendency for particles near the edge of the emittance to lose transverse energy and fall back towards the centre. In an electron machine the emittance, determined by the equilibrium between these two effects, grows with  $E^2$ . Liouville's theorem only applies to particles guided by external fields and does not apply to electron machines where particles emit some of their own energy.

## 2.3 Definitions of emittance

Within the boundary of transverse phase space, the emittance of the beam, particles are usually distributed in a population which appears Gaussian when projected on a vertical or horizontal plane. The emittance boundary is traditionally chosen to include 95% of a Gaussian beam or  $\approx 2 \sigma_y$  in a proton machine. In an electron machine a  $2 \sigma$  boundary would be too close to the beam. An aperture stop placed at this distance would rather rapidly absorb

most of the beam as quantum emission stirs it up in phase space. The physical boundary allowed, depends on the lifetime required but is in the region  $6\sigma$  to  $10\sigma$ . What is normally quoted as an emittance for an electron beam corresponds to  $\sigma$  of the Gaussian projection. There is consequently a factor 4 between emittance defined by electron and proton people. Beware! Just to complicate matters proton people have of late taken to using the  $\sigma^2/\beta$  definition of emittance used by electron people.

## 2.4 Acceptance

In contrast, the acceptance,  $A$ , the size of the hole which is the vacuum chamber transformed into phase space, is unambiguous. Remembering from the lectures on focusing that particles follow an elliptical trajectory with axes

$$\Delta y = \pm\sqrt{\epsilon\beta}, \quad \Delta y' = \pm\sqrt{\epsilon/\beta}$$

and area  $\pi\epsilon$  we can write

$$\begin{aligned}\epsilon_{\text{protons}} &= \frac{(2\sigma)^2}{\beta} \\ \epsilon_{\text{electrons}} &= \frac{\sigma^2}{\beta} \\ A &= \frac{r^2}{\beta}\end{aligned}$$

where  $r$  is the semi-axis of the chamber.

## 2.5 Measurement of emittance

It is my intention to give you some idea of the means used to measure the various optical quantities as they are mentioned. The methods of observing particle beams are much more limited than theory might suggest and it is as well to bear this in mind and aim for simplicity in arriving at a design rather than to rely upon a complex procedure.

The most reliable and straightforward way to measure proton beam size is to drive a scraper into the beam or move the beam across the scraper and integrate the beam loss curve. Electron machines observe beam size by refocusing the synchrotron light emitted, though the beam dimensions can be so small that special image scanning techniques are needed to resolve the image. A clever, non-destructive device used for proton machines is the Ionisation Beam Scanner [6]. Figure 7 shows the principle in which a zero electrical potential scans across the beam allowing electrons to be collected from its surface as the beam ionises the residual gas. A typical signal is given in Fig. 8.

This IBS has the advantage that it can give a "mountain range" display of how the beam shrinks during acceleration (Fig. 9). Its disadvantage is that the space charge fields of an intense beam can distort the width of the peak which represents the beam. For this reason it is rarely used nowadays except, as in this talk, for pedagogical reasons.

Another diagnostic tool is a wire which is scanned across the beam very rapidly. Secondary particles generated are counted with a scintillation telescope. Here one must be careful not to dilate the beam by scattering (or burn the wire).

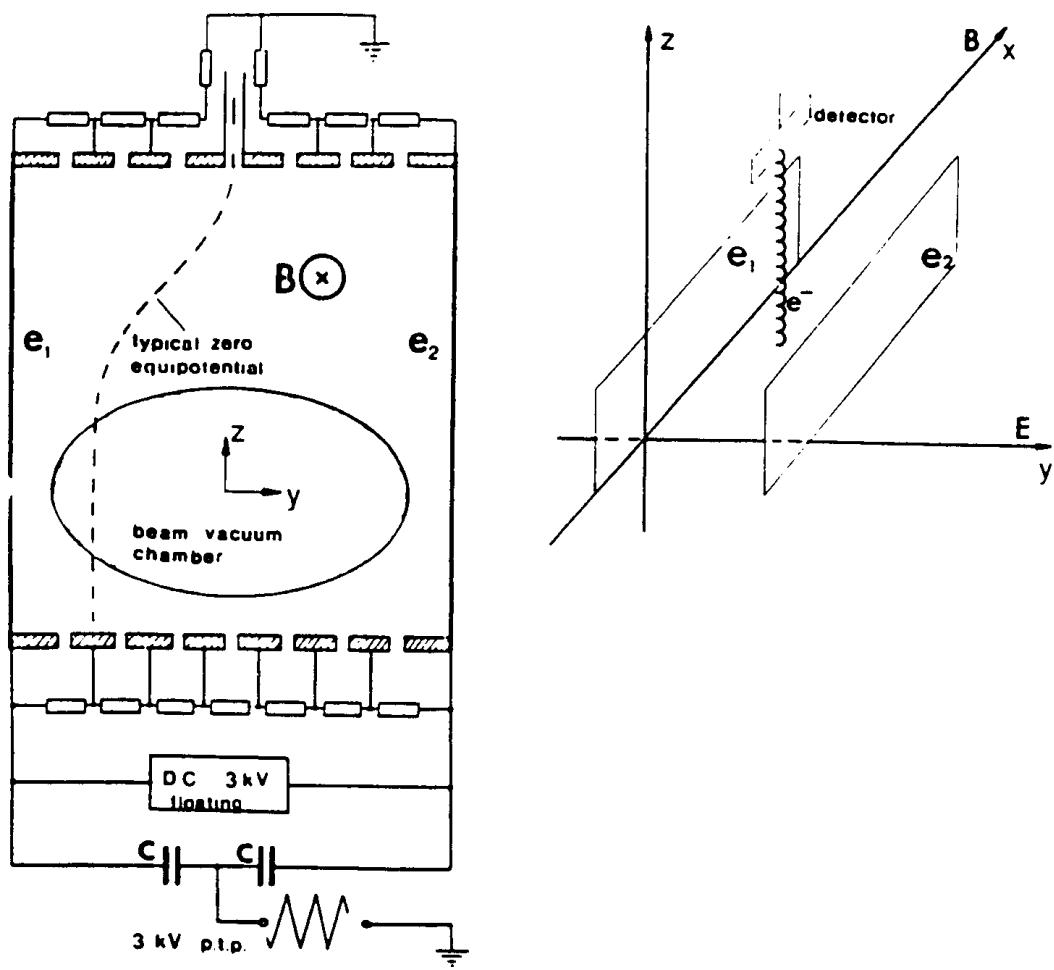


Fig. 7 Diagram showing IBS principle of beam scanning

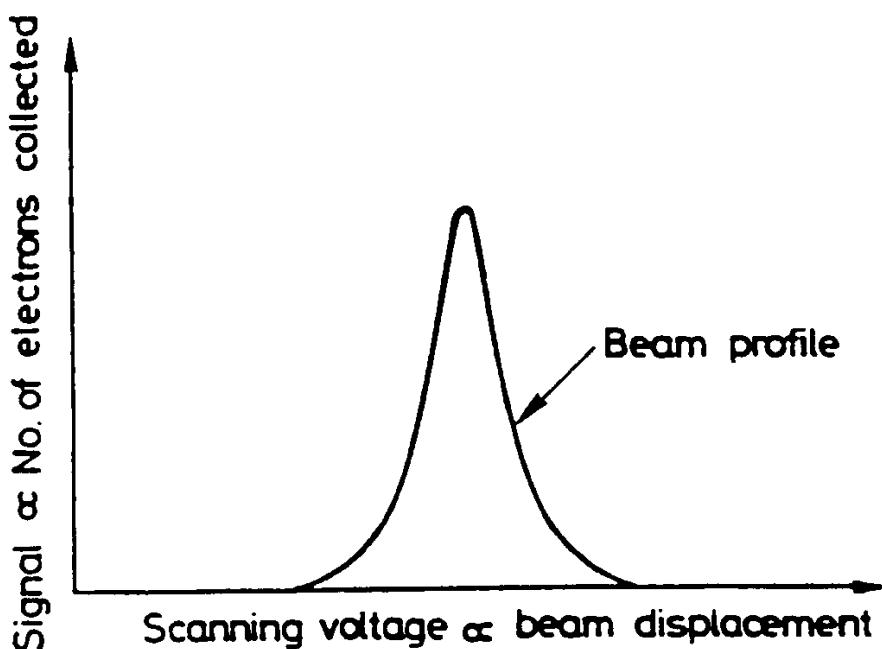


Fig. 8 Ideal ion beam scanner scope trace

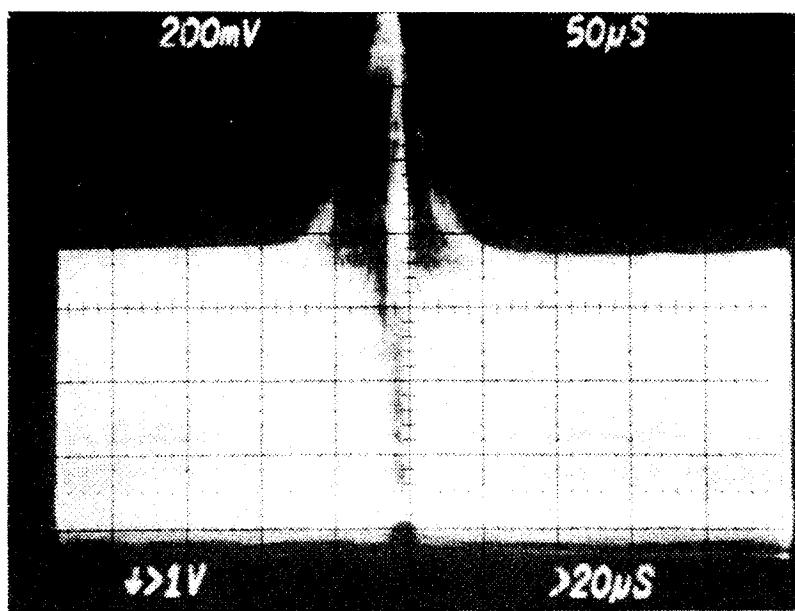


Fig. 9 Mountain range display and horizontal profile; 1 cm/division horizontal profile at 350 GeV

### 3. A SIMPLIFIED TREATMENT OF BETATRON MOTION

The earlier lectures on focusing treat betatron motion with all the rigour necessary to design a machine. However some readers who are new to the field might find the following sections too confusing if we carry through all the terms from the rigorous theory. We shall therefore use two models of the motion which are approximate but graphic. We shall confine our examination of the motion to maxima or minima in the beta function where the ellipse is upright as in Fig. 10.

In Fig. 10b we see the ellipse, plotted at a point where the amplitude function  $\beta$  is large (as it is near F quadrupoles). The ellipse will be very wide and not very high in divergence angle and it is in such positions that a small angular kick has the greatest effect on the beam. Imagine, for example, how little angular displacement is needed to move the ellipse by its own height and increase the emittance by a factor 2. Thus, in a machine with a lattice like Fig. 1 most of the damage done to the beam by bad fields which cause angular kicks, happens near the F quadrupoles.

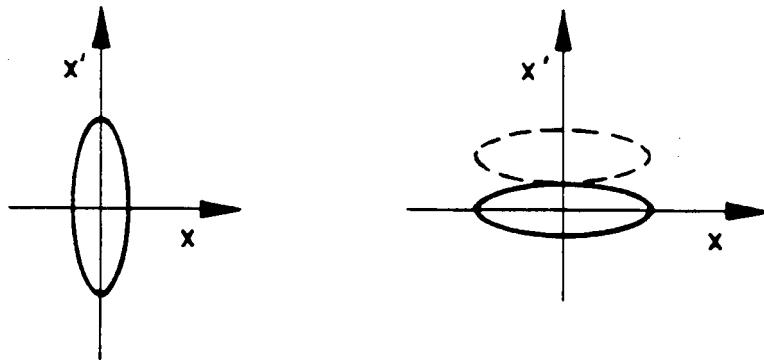


Fig. 10 Phase-space diagram at (a) a  $\beta$  minimum and (b) a  $\beta$  maximum

### 3.1 Circle approximation

So predominant is the effect of perturbations near  $\hat{\beta}$  positions that you can often do quite good "back of the envelope" calculations by closing your eyes to what happens to the protons in between F quadrupoles. At F quadrupoles the ellipse always looks the same, i.e. upright, with semi-axes in displacement and divergence

$$\sqrt{(\beta\epsilon)}, \sqrt{(\epsilon/\beta)}$$

This can be reduced to a circle radius  $\sqrt{\beta\epsilon}$  by using the new coordinates

$$\begin{aligned} y &= y \\ p &= \beta y' . \end{aligned}$$

If the machine has 108 periods and a  $Q$  of 27.6, the proton advances in phase by  $2\pi Q/108$  from one period to the next; this is just the angle subtended at the centre of the circle multiplied by  $Q$ . After one turn of the machine, it has made 27 revolutions of the circle plus an angle of  $2\pi$  multiplied by the fractional part of  $Q$ , see Fig. 11.

This renormalization of the phase space can be done in a more rigorous way by choosing new variables  $(\eta, \psi)$  which transform the distortion of the phase and amplitude so that the motion becomes that of a harmonic oscillator. We must, of course, transform back again to see physical displacements, but the mathematics becomes more transparent. This simplification is discussed in the next section.

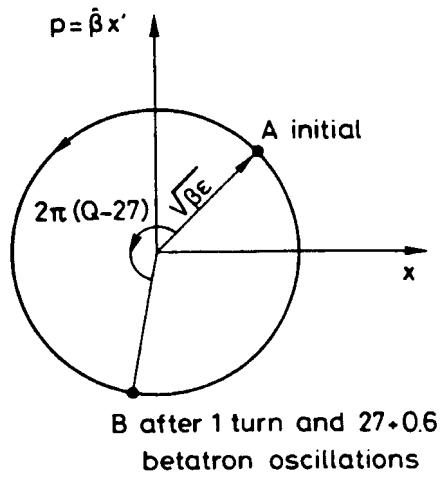


Fig. 11 Circle diagram (locus at F quadrupoles)

### 3.2 The $(\eta, \psi)$ description of AG focusing

A more rigorous renormalization of phase space which does not imply any approximation but which simplifies the problem is the  $(\eta, \psi)$  transformation to convert Hill's equation into that of a harmonic oscillator:

$$\frac{d^2\eta}{d\psi^2} + Q^2\eta = g(\psi)$$

where  $g(\psi)$  is the azimuthal pattern of some perturbation of the guide field related to

$$F(s) = \frac{\Delta B(s)}{B\rho}.$$

In the ideal case  $g(\psi)$  is everywhere zero.

I will not bother you with how this transformation is found, but just state it. The new coordinates are related to the old:

$$\begin{aligned}\eta &= \beta^{-1/2}y \\ \psi &= \int \frac{ds}{Q\beta}, \quad g(\psi) = Q^2\beta^{3/2}F(s),\end{aligned}$$

where  $\psi$  advances by  $2\pi$  every revolution. It coincides with  $\theta$  at each  $\beta$  or  $\beta'$  location and does not depart very much from  $\theta$  in between.

## 4. THE $Q$ VALUES

A full understanding of transverse dynamics is rather difficult until one has an almost tactile appreciation of the nature of the oscillations. The simple models of the last section are intended to help but even more insight may be had from contemplating the methods of measuring the  $Q$  values. A number of methods are possible and each reveals a different aspect of the motion.

## 4.1 Measurement of $Q$ by kicking

The method is to fire a kicker magnet with a pulse lasting less than one turn and observe the way in which the centre of charge of the beam oscillates as it passes a pick-up on sequential turns (see Fig. 12).

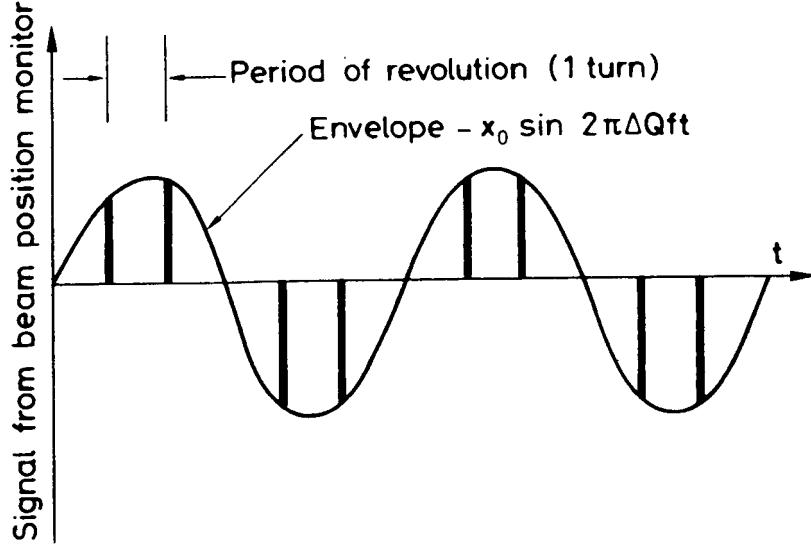


Fig. 12 Ideal  $Q$ -measurement signal following a kick which excites coherent betatron motion.  $Q = \text{integer } \pm 1/6$ .

In order to understand Fig. 12 it is convenient to imagine a beam consisting of one short longitudinal bunch. The line density current passing a detector is then a Fourier series

$$\rho(t) = \sum_n a_n \sin 2\pi n f_0 t .$$

This beam position detector sees the betatron oscillations following the kick as:

$$y(t) = y_0 \cos 2\pi f_0 Q t$$

but modulated by  $\rho(t)$ .

An oscilloscope connected to the pick-up will give a display of the modulated signal, the product of  $\rho$  and  $y$

$$\rho(t)y(t) = \frac{1}{2} \sum_n a_n y_0 [\sin 2\pi(n+Q)f_0 t + \sin 2\pi(n-Q)f_0 t] .$$

The envelope of the oscilloscope signal will be the slowest of these terms in which  $(n-Q)$  is the fractional part of  $Q$ . The other terms in the series reconstruct the spikes in the signal occurring once per turn. Note that I have made use of the elementary relation:

$$\sin a \cos b = \frac{1}{2} [\sin(a+b) + \sin(a-b)] .$$

This will be often used when we come to look at resonances.

## 4.2 R.F. knockout and $Q$ measurement

A very simple form of resonance can be induced by applying a deflecting field with the frequency of a betatron sideband as the frequencies  $(n \pm Q)f_0$  are called. We can easily invert the above treatment to show that if you apply a signal of the form

$$\sin[2\pi(n - Q)f_0 t]$$

then the particle passing the electrodes of the deflector sees the deflector every turn and experiences a kick

$$\sin[2\pi n f_0 t] \sin[2\pi(n - Q)n f_0 t].$$

Using again the elementary relation of Section 4.1 we see that the  $(a - b)$  component is in resonance with the betatron motion

$$\sin 2\pi Q f_0 t.$$

Once in resonance, the particle is deflected on each turn by a kick which increases its amplitude of transverse oscillations in phase with the excitation and blows up the beam. Note that this like the previous method gives a value of the fractional part of  $Q$  with respect to the nearest integer but gives no information about which integer this is.

## 4.3 Measurement by analysing the frequencies emitted by the beam

It is also possible nowadays to detect these betatron frequencies among the statistical noise signal detected by a transverse pick-up which can be just a pair of plates and displayed with an integrating spectrum analyser which is really a scanning radio receiver connected to an oscilloscope. Peaks appear as sidebands to the revolution frequency in the display of response versus frequency (Fig. 13). Their separation is  $2\Delta Q f_0$  where  $\Delta Q$  is the fractional part of  $Q$ .

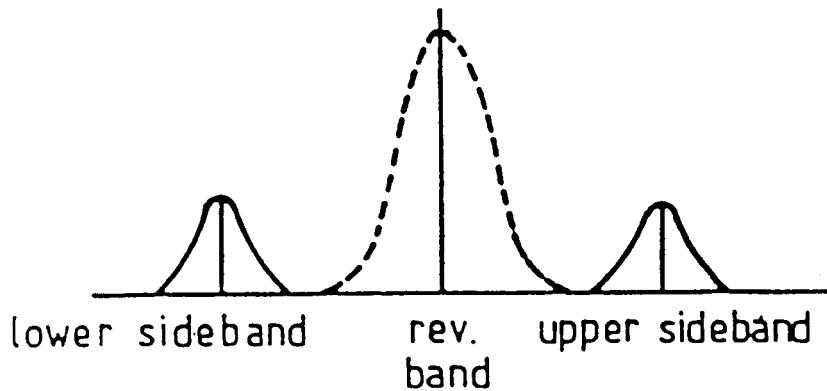


Fig. 13 Transverse pickup seen on a spectrum analyser

One may wonder why the particles, evenly spread around the ellipse in phase space, can generate such a signal. The answer is that they are finite in number and the pick-up samples

but a small fraction of them. In a sample there are always significant statistical fluctuations of the centre of charge, or mean displacement, which the spectrum analyser picks up.

## 5. CLOSED-ORBIT DISTORTION

As an illustration of the power of  $(\eta, \psi)$  coordinates we look at closed-orbit distortions which are produced by field errors. Even the best synchrotron magnets cannot be made absolutely identical. Each magnet differs from the mean by some small error in integrated strength:

$$\delta(B\ell) = \int B \, d\ell - \left( \int B \, d\ell \right)_{\text{ideal}} .$$

These and other machine imperfections, such as survey errors which can be expressed as equivalent to field errors, are randomly spread around the ring.

We can use the  $(\eta, \psi)$  coordinates to find out how this perturbs a proton which would otherwise have had zero betatron amplitude. Such a proton no longer goes straight down the centre of the vacuum chamber but follows a perturbed closed orbit about which the normal betatron motion of the other protons can be superimposed.

One of the most important considerations in designing a machine is to keep this closed orbit distortion to a minimum because it eats up available machine aperture. Also, once we have succeeded in getting a few turns round the machine, we want to reduce this distortion with correcting dipole magnets. As a first step let us consider the effect on the orbit of such a correcting dipole located at a position where  $\beta = \beta_K$  and observed at another position.

A short dipole (we shall assume it is a delta function in  $s$ ) makes a constant angular kick in divergence

$$\delta y' = \delta(B\ell) / (B\rho) ,$$

which perturbs the orbit trajectory which elsewhere obeys

$$\frac{d^2\eta}{d\psi^2} + Q^2\eta = 0$$

$$\eta = \eta_0 \cos(Q\psi + \lambda) .$$

We choose  $\psi = 0$  origin to be diametrically opposite the kick. Then by symmetry  $\lambda = 0$  and the "orbit" is that shown in Fig. 14.

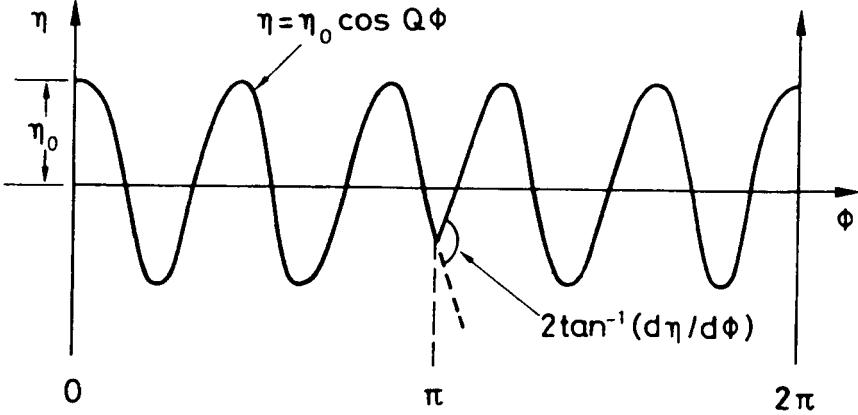


Fig. 14 Tracing the closed orbit for one turn in  $(\eta, \psi)$  space with a single kick at  $\psi = \pi$ . (The  $Q$  value is about 5.6.)

Since, by definition, the trajectory is closed, continuity demands that the kick  $\delta y'$  matches the change in slope at  $\psi = \pi$ , the location of the dipole.

Differentiating the orbit equation

$$\frac{d\eta}{d\psi} = -\eta_0 Q \sin Q\psi = -\eta_0 Q \sin Q\pi, \quad \text{at } \psi = \pi.$$

To relate this to the real kick we use

$$\frac{d\psi}{ds} = \frac{1}{Q\beta_K}, \quad \frac{dy}{ds} = \sqrt{\beta_K} \frac{d\eta}{ds},$$

therefore

$$\begin{aligned} \frac{\delta y'}{2} &= \frac{\delta(B\ell)}{2B\rho} = \frac{dy}{ds} = \sqrt{\beta_K} \frac{d\eta}{d\psi} \frac{d\psi}{ds} = \frac{-\eta_0}{\sqrt{\beta_K}} \sin \pi Q \\ \eta_0 &= \frac{\sqrt{\beta_K}}{2|\sin \pi Q|} \delta y'. \end{aligned}$$

Returning to physical coordinates we can write the orbit's equation in the range  $-\pi < \psi(s) < \pi$ :

$$y = \sqrt{\beta(s)} \eta_0 \cos Q\psi(s) = \left[ \frac{\sqrt{\beta(s)\beta_K}}{2|\sin \pi Q|} \frac{\delta(B\ell)}{B\rho} \right] \cos Q\psi(s).$$

The expression in square brackets is the maximum amplitude of the perturbation at  $\beta(s)$ .

The above expression is rigorous but as an example of the use of the circle approximation consider the special case where the kink and observation are at the same value of beta. We see quite clearly from Fig. 15 how the equation for the amplitude of the distortion appear.

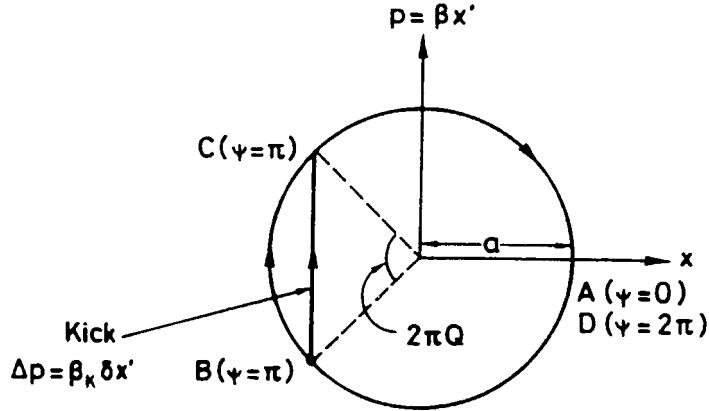


Fig. 15 Tracing a closed orbit for one turn in the circle diagram with a single kick. The path is ABCD.

In estimating the effect of a random distribution of dipole errors we must take the r.m.s. average, weighted according to the  $\beta_K$  values over all of the kicks  $\delta y_i$  from the N magnets in the ring. The expectation value of the amplitude is:

$$\begin{aligned} \langle y(s) \rangle &= \frac{\sqrt{\beta(s)}}{2\sqrt{2} \sin \pi Q} \sqrt{\sum_i \beta_i \delta y_i^2} \\ &\approx \frac{\sqrt{\beta(s)\beta}}{2\sqrt{2} \sin \pi Q} \sqrt{N} \frac{(\Delta B\ell)_{rms}}{B\rho}. \end{aligned}$$

The factor  $\sqrt{2}$  comes from averaging over all the phases of distortion produced.

The principal imperfections in a synchrotron causing orbit distortion are shown in Table 1 [7]. The first line in the table represents the random variations in the position of quadrupole magnets with respect to their ideal location. A small displacement of a quadrupole gives an effective dipole perturbation,  $k_l \Delta y$ . The tilt of bending magnets causes a small resultant dipole in the horizontal direction which deflects vertically. Obviously random errors in magnet gap, length or in the coercivity of the steel yoke which determines remanent field contribute to the third line. Both remanent and stray fields in straight sections tend to be constant and their effect scales as  $1/B$  as the machine pulses. Their effect should therefore be evaluated where it is worst, i.e. at injection. In a modern superconducting machine the persistent current fields play the role of remanent effects.

**Table 1**  
Sources of Closed Orbit Distortion

Type of element	Source of kick	r.m.s. value	$\langle \Delta B\ell/(B\rho) \rangle_{rms}$	plane
Gradient magnet	Displacement	$\langle \Delta y \rangle$	$k_l i \langle \Delta y \rangle$	$x, z$
Bending magnet (bending angle = $\theta_i$ )	Tilt	$\langle \Delta \rangle$	$\theta_i \langle \Delta \rangle$	$z$

Bending magnet	Field error	$\langle \Delta B/B \rangle$	$\theta_i \langle \Delta B/B \rangle$	$x$
Straight sections (length = $d_i$ )	Stray field	$\langle \Delta B_s \rangle$	$d_i \langle \Delta B_s \rangle / (B\rho)_{inj}$	$x, z$

In designing a machine it used to be conventional wisdom to make sure that the vacuum chamber will accommodate twice this expectation value. The probability of no particles making the first turn is thus reduced to a mere 2%. More modern designs rely on closed-orbit steering to thread the first turn and thereafter assume that orbit correction to a millimeter or so will be feasible.

## 5.1 The Fourier harmonics of the error distribution

One of the advantages of reducing the problem to that of a harmonic oscillator in  $(\eta, \psi)$  coordinates is that perturbations can be treated as the driving term of the oscillator, broken down into their Fourier components, and the whole problem solved like the forced oscillations of a pendulum. The driving term is put on the right hand side of Hill's equation:

$$\frac{d^2\eta}{d\psi^2} + Q^2\eta = Q^2 \sum_{n=1}^{\infty} f_k e^{ik\psi} = Q^2 \beta^{3/2} F(s) ,$$

where  $F(s)$  is the azimuthal pattern of the perturbation  $\Delta B/(B\rho)$ ; and  $Q^2\beta^{3/2}$  comes from the transformation from physical coordinates to  $(\eta, \psi)$ .

The Fourier amplitudes are defined:

$$f(\psi) = \beta^{3/2} F(s) = \sum_k f_k e^{ik\psi} ,$$

where

$$f_k = \frac{1}{2\pi} \int_0^{2\pi} f(\psi) e^{-ik\psi} d\psi = \frac{1}{2\pi Q} \oint \beta^{1/2} F(s) e^{-ik\psi} ds .$$

We can then solve Hill's equation as

$$\eta = \sum_{n=1}^{\infty} \frac{Q^2 f_k}{Q^2 - k^2} e^{ik\psi} \quad (\text{or its real part}) .$$

But be careful. Before doing the Fourier analysis,  $\Delta B$  must be multiplied by  $\beta^{1/2}$  if the physical variable  $s$  is chosen as an independent variable, or  $\beta^{3/2}$  if  $\psi$ , the transformed phase, is used.

Looking carefully at the above expression, we see that this differs from the general solutions

$$\eta = \eta_0 e^{\pm iQ\psi}$$

which describe betatron motion about the equilibrium orbit, because the wave number is an integer  $k$ . In fact it is a *closed* orbit, a particular solution of Hill's differential equation, to which we must add the general solutions which describe betatron oscillations about this orbit.

The function  $Q^2/(Q^2 - k^2)$  is sometimes called *the magnification factor* for a particular Fourier component of  $\Delta B$ . It rises steeply when the wave number  $k$  is close to  $Q$ , and the effect of the two Fourier components in the random error pattern with  $k$  values adjacent to  $Q$  accounts for about 60% of the total distortion due to all random errors. Figure 16 shows a closed orbit pattern from electrostatic pick-ups in the FNAL ring, whose  $Q$  is between 19 and 20. The pattern shows strong components with these wave numbers. If  $Q$  is deliberately tuned to an integer  $k$ , the magnification factor is infinite and errors of that frequency make the proton walk out of the machine. This is in fact an integer resonance driven by dipole errors.

One instructive method of correcting orbit distortion is to apply a pattern of dipole correctors which excite an equal and opposite Fourier component of error at an integer close to  $Q$ .

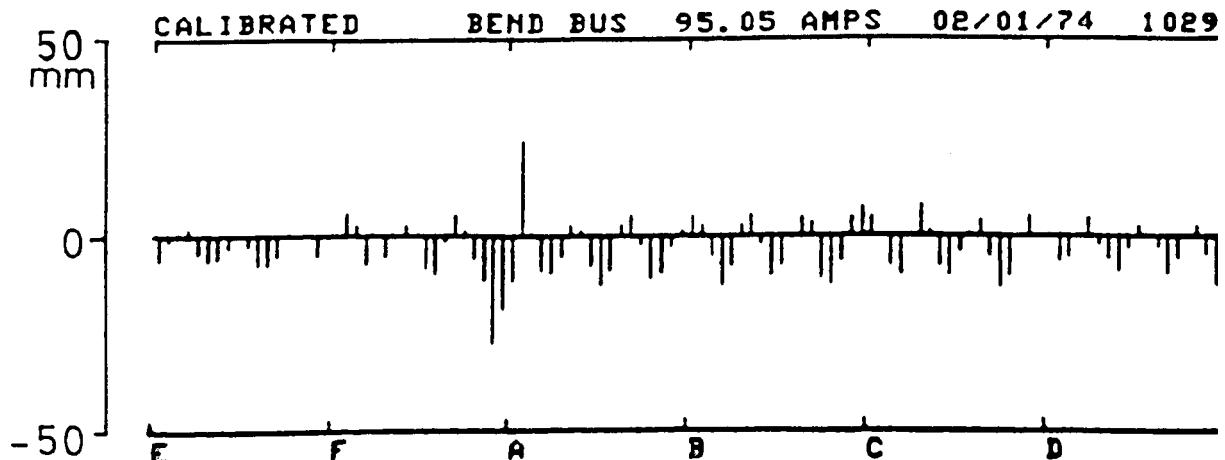


Fig. 16 FNAL main ring electrostatic pick-ups show closed orbit around the ring ( $Q \approx 19.2$ )

## 5.2 Closed-orbit bumps

It is often important to deliberately distort a closed orbit bump at one part of the circumference without affecting the central orbit elsewhere. A typical example of this is to make the beam ride close to an extraction septum or within the narrow jaws of an extraction kicker magnet placed just outside the normal acceptance of the ring.

If one is lucky enough to find two positions for small dipole deflecting magnets spaced by  $\pi$  in betatron phase and centred about the place where the bump is required, the solution is very simple. The distortion produced is:

$$y(s) = \delta \sqrt{\beta(s)\beta_k} \sin(\phi - \phi_0)$$

where  $\beta(s)$  is the beta function at  $s$ ,  $\beta_k$  is the beta function at the deflector and

$$\delta = \frac{\Delta(B\ell)}{B\rho} .$$

This half-wave bump has a very simple configuration in normalised phase space (Fig. 17) We can immediately see that the central orbit (at the origin) is not disturbed elsewhere).

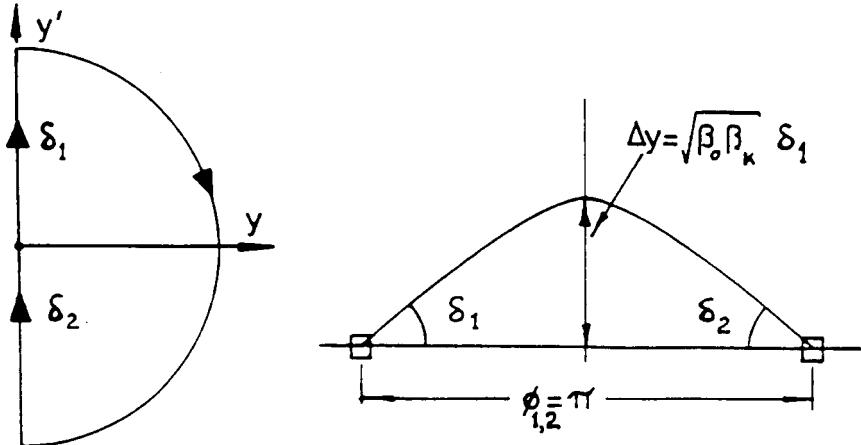


Fig. 17 An exact half-wave bump using two dipoles

Note that the magnitude of the bump is not only proportional to the root of the local  $\beta(s)$  but is proportional to  $\sqrt{\beta_k}$ . Since  $\beta$  is largest at F quadrupoles this is clearly where one should locate dipole bumpers.

Very often F-quadrupoles are not  $\pi$  apart in phase but  $2\pi/3$ . this means the dipoles are slightly less effective ( $\sin \pi/3 = \sqrt{3}/2$ ) but also we must introduce a third dipole to form a triad. The third dipole is best located near the peak of the bump.

Figure 18 shows how the three bumps add up in normalised phase space. The case illustrated is the general one with dipoles of different strengths and spaced differently in phase.

In order to find an exact solution to the problem of a triad bump we use the matrix which transforms a point in phase space from one location to another (Eq. 1 of Ref. [2]).

$$\begin{pmatrix} y \\ y' \end{pmatrix} = \begin{pmatrix} (\sqrt{\beta} / \sqrt{\beta_0})(\cos \Delta\phi + \alpha_0 \sin \Delta\phi) & , \sqrt{\beta_0 \beta} \sin \Delta\phi \\ (-1 / \sqrt{\beta_0 \beta})\{(\alpha - \alpha_0) \cos \Delta\phi + (1 + \alpha \alpha_0) \sin \Delta\phi\} & , (\sqrt{\beta} / \sqrt{\beta_0})(\cos \Delta\phi - \alpha \sin \Delta\phi) \end{pmatrix} \begin{pmatrix} y_0 \\ y'_0 \end{pmatrix}$$

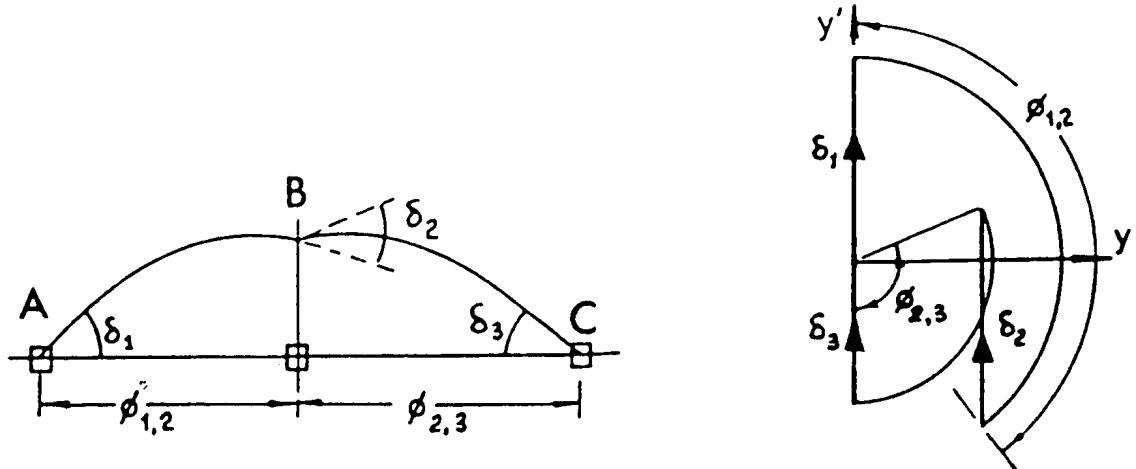


Fig. 18 A general bump using three dipoles

It is the element which links  $y$  to  $y'$  which describes the trajectory. Following the kick  $\delta_i$

$$y_2 = \delta_1 \sqrt{\beta_2 \beta_1} \sin(\phi_2 - \phi_1) .$$

The same argument can be used to describe the trajectory working back from  $\delta_3$

$$y_2 = -\delta_3 \sqrt{\beta_2 \beta_3} \sin(\phi_2 - \phi_3) .$$

The kick  $\delta_2$  must be the change in the derivative

$$\delta_2 = \delta_1 \frac{\sqrt{\beta_1}}{\sqrt{\beta_2}} \cos(\phi_2 - \phi_1) + \delta_3 \frac{\sqrt{\beta_3}}{\sqrt{\beta_2}} \cos(\phi_3 - \phi_2) .$$

Here the  $\sqrt{\beta_2}$  in the denominator comes from multiplying by  $d\phi/ds = 1/\beta_2$  at the location of the cusp.

We can rewrite these relations

$$\begin{aligned} \sqrt{\beta_1} \delta_1 \sin \phi_{1,2} &= \sqrt{\beta_3} \delta_3 \sin \phi_{2,3} \\ \sqrt{\beta_2} \delta_2 &= \sqrt{\beta_1} \delta_1 \cos \phi_{1,2} + \sqrt{\beta_3} \delta_3 \cos \phi_{2,3} . \end{aligned}$$

These relations can be seen to be true for the triangle of sides,  $\delta / \sqrt{\beta}$  and angles,  $\phi$  which can be solved by the well-known symmetric relation

$$\frac{\delta_1 \sqrt{\beta_1}}{\sin \phi_{2,3}} = \frac{\delta_2 \sqrt{\beta_2}}{\sin \phi_{3,1}} = \frac{\delta_3 \sqrt{\beta_3}}{\sin \phi_{1,2}} .$$

### 5.3 The measurement and correction of closed orbits

Electrostatic plates with diagonal slots are commonly used to measure the transverse position of a bunched beam. We have seen that the predominant harmonic in the uncorrected orbit is close to  $Q$  and to establish its amplitude and phase one really needs four pick-ups per

wavelength. Given the present fashion for FODO lattices with about  $90^\circ$  per cell and the need to measure in both planes, the final solution is usually one pickup at each quadrupole. The ones at F-quadrupoles, where beta is large horizontally, are the most accurate for the horizontal plane while others at D-quadrupoles are best for the vertical correction.

Similar arguments lead us to have one horizontally deflecting dipole at each F-quadrupole where  $\beta_x$  is large, and one vertically deflecting dipole at each D-quadrupole, where  $\beta_z$  is large.

Clearly so many correcting dipoles are unnecessary if only the two principal harmonics are to be corrected. In Fig. 19, one can excite the 19th harmonic with a single dipole, though a pair in opposition are necessary if their bending effect is to cancel and a further pair are needed in quadrature for the phase making four per harmonic and per plane. However, studies show that this simple harmonic correction still leaves about 30% of the initial distortion.

Another method consists in applying a set of superposed beam bumps formed along the lines calculated above; each a triad which compensates the measured orbit position at its centre. Given the power of modern computers this kind of correction can be calculated and applied all round the ring in a few seconds.

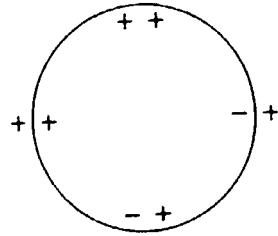


Fig. 19 Diagram showing the sign of correction dipoles necessary to excite or compensate even or odd Fourier components of distortion around the ring of a synchrotron

Some machines do not have dipole correctors which are sufficiently strong to correct an orbit at their top energy and quadrupole magnets must be displaced upwards or sideways a distance  $\Delta y$  to apply an effective dipole:

$$\frac{\Delta(B\ell)}{B\rho} = k\Delta y .$$

Such displacements are tedious to apply to all quadrupoles in such large machines as the SPS or LEP and indeed the accumulated effect of errors in moving so many quadrupoles by a few *tenths* of a millimetre might even make the orbit worse.

An alternative [8] to moving all quadrupoles (or powering all dipoles) is to select those which are most effective in correcting the orbit. Stored in the computer is a large matrix  $\mathbf{G}$  with as many rows as pick-ups and as many columns as correctors. Each term describes the effect  $Y_i$  of a corrector  $\Delta_j$  at the  $i$ th pick-up

$$\mathbf{Y} = \mathbf{G}\Delta .$$

Once the orbit is measured the column matrix  $\mathbf{Y}$  is entered and can be converted into a set of corrections  $\Delta$ :

$$\Delta = \left\{ -(\tilde{\mathbf{G}}\mathbf{G})^{-1}\tilde{\mathbf{G}} \right\} \mathbf{Y}$$

But since we first want to select the most effective correctors, we examine the terms of  $\tilde{\mathbf{Y}}\mathbf{G}\Delta$  which is a measure of efficiency. The best corrector is retained while a search is made for another one to form a doublet. At the cost of considerable computing time one can find any number of best correctors. The method also has the advantage that it can be used if one or more of the pick-ups is out of action.

One last comment on orbit correction is that while originally invented to save on magnet aperture, orbit corrections now seem an essential procedure to reduce the effects of non-linearities in the dynamics of synchrotrons. One example is that orbit distortion in the sextupoles, which all machines have for correcting chromaticity, will generate a pattern of quadrupole gradient errors and drive half-integer stopbands.

## 6. GRADIENT ERRORS

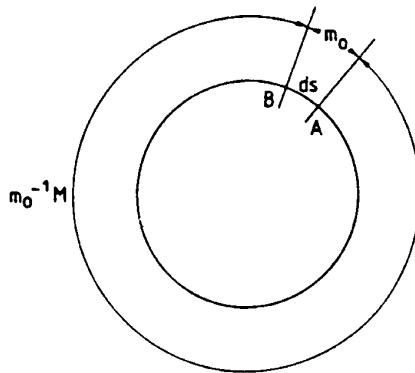


Fig. 20 Matrix representation of a small quadrupole,  $m_0$  subject to an error which is a component of the matrix for the whole ring,  $M$

Quadrupoles as well as bending magnets may have errors. Understanding the effect of such gradient errors is a useful preparation for the study of non-linear errors. We represent a ring of magnets as a circle in Fig. 20 and the matrix for one turn starting at A as:

$$M_0(s) = \begin{pmatrix} \cos \varphi_0 + \alpha_0 & \sin \varphi_0 , & \beta_0 & \sin \varphi_0 \\ -\gamma_0 & \sin \varphi_0 , & \cos \varphi_0 - \alpha_0 & \sin \varphi_0 \end{pmatrix}.$$

Now consider a small gradient error which afflicts a quadrupole in the lattice between B and A. The unperturbed matrix for this quadrupole is  $m_0$  and when perturbed the quadrupole matrix is

$$m_0 = \begin{pmatrix} 1 & 0 \\ -k_0(s)ds & 1 \end{pmatrix},$$

and perturbed, a matrix

$$m = \begin{pmatrix} 1 & 0 \\ -[k_0(s) + \delta k(s)]ds & 1 \end{pmatrix}.$$

The unperturbed transfer matrix for the whole machine:

$$M_0(s) = \begin{pmatrix} \cos\phi_0 + \alpha_0 & \sin\phi_0 & \beta_0 & \sin\phi_0 \\ -\gamma_0 & \sin\phi_0 & \cos\phi_0 - \alpha_0 & \sin\phi_0 \end{pmatrix}$$

includes  $m_0$ .

To find the perturbed transfer matrix we make a turn, back-track through the small unperturbed quadrupole ( $m_0^{-1}$ ), and then proceed through the perturbed quadrupole ( $m$ ). Translated into matrix algebra,

$$M(s) = mm_0^{-1}M_0 .$$

Now

$$mm_0^{-1} = \begin{pmatrix} 1 & 0 \\ -\delta k(s_1)ds & 1 \end{pmatrix} .$$

So

$$M = \begin{pmatrix} \cos\phi_0 + \alpha_0 \sin\phi_0 & \beta_0 \sin\phi_0 \\ -\delta k(s)ds(\cos\phi_0 + \alpha_0 \sin\phi_0) - \gamma \sin\phi_0 & -\delta k(s)ds\beta_0 \sin\phi_0 + \cos\phi_0 - \alpha_0 \sin\phi_0 \end{pmatrix} .$$

Now  $1/2(Tr M) = \cos\phi$ . So the change in  $\cos\phi$  is

$$\Delta(\cos\phi) = -\Delta\phi \sin\phi_0 = \frac{\sin\phi_0}{2} \beta_0(s) \delta k(s) ds$$

$$2\pi\Delta Q = \Delta\phi = \frac{\beta(s)\delta k(s)ds}{2} .$$

Since betatron phase is not involved in this equation we are tempted to integrate around the ring to obtain

$$\Delta Q = \frac{1}{4\pi} \int \beta(s) \delta k(s) ds .$$

This result is often used in accelerator theory and is surprising in that we see that the change is independent of the phase of the perturbation. However, this equation is only approximately true since as we add each elemental focusing error it modifies  $\beta(s)$  as well as  $Q$  so that there is a higher-order term which should be included if one wants accurate numerical results [see Ref. [9] Eqs. (4.32) to (4.37)]. Nevertheless, used with discretion it is sufficiently accurate to explain the physical basis of the resonant phenomena we shall be discussing in later sections which can usually only be estimated to within a factor of 2 anyway.

The reason for our concern about the change in tune or phase advance which results from errors is that we must steer  $Q$  well away from certain fractional values which can cause the motion to resonate and result in loss of the beam. To understand how some  $Q$  values are dangerous let us return to the case of closed orbit distortion. Earlier we found the orbit distortion amplitude:

$$\hat{y} = \frac{\sqrt{\beta\beta_\kappa}}{2 \sin \pi Q} \cdot \frac{\delta(B\ell)}{B\rho} .$$

Clearly this will become infinite if  $Q$  is an integer value. What happens physically is that the beam receives a kick at the same phase on every turn and just spirals outwards. An error in gradient can have the same effect if the  $Q$  value is close to one of the lines:

$$\left. \begin{array}{l} 2Q_h = p , \quad 2Q_v = p \\ Q_h = Q_v = p , \quad Q_h + Q_v = p \end{array} \right\}$$

where  $p$  is an integer.

At this stage in the description of transverse dynamics we can only hint at the explanation for this. Particles spiral outwards in phase space if the perturbation has the same effect on each turn. The perturbation from a dipole is independent of the transverse displacement and can only build up in this way if the particle returns to the same point in phase space on each turn ( $Q = p$ ). A quadrupole error has field proportional to  $x$  and if a particle makes half turns in phase space it will see alternately positive and negative kicks in divergence but both will reinforce the growth. One may extend this argument to understand why sextupole errors which have a quadratic  $x$  dependence excite third-integer "resonances" as they are called near the lines.

$$\left. \begin{array}{l} 3Q_h = p \\ 2Q_h + Q_v = p \\ Q_h + 2Q_v = p \\ 3Q_v = p \\ 2Q_h - Q_v = p \\ Q_h - 2Q_v = p \end{array} \right\}$$

## 7. THE WORKING DIAGRAM

This is simply a diagram with  $Q_H$  and  $Q_V$  as its axes. The beam can be plotted on it as a point but because there is a certain  $Q$ -spread among protons of different momenta it is better to give the point a finite radius  $\Delta Q$  (Fig. 21).

We plot on the diagram a mesh of lines which mark danger zones for the protons. We have hinted that if  $Q$  in either the vertical or the horizontal plane is a simple vulgar fraction, then

$$nQ = p ,$$

where  $n$  and  $p$  are integer and  $n < 5$ , a resonance takes over and walks the proton out of the beam. In general this is true when

$$\ell Q_H + m Q_V = p ,$$

where  $|\ell| + |m|$  is the order of the resonance and  $p$  is the azimuthal frequency which drives it.

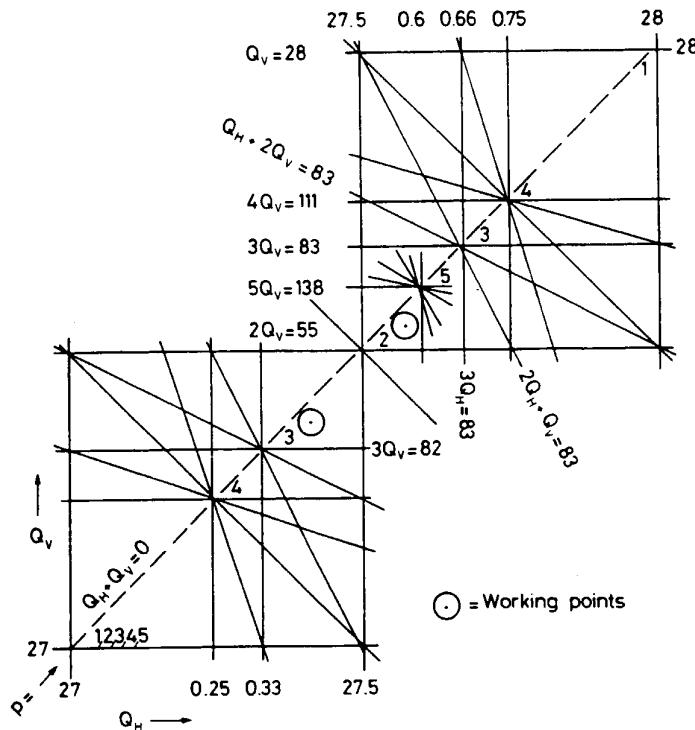


Fig. 21 SPS working diamond

This equation just defines a set of lines in the  $Q$  diagram for each order of resonance and for each value of the integer  $p$ . Figure 21 shows these lines for the SPS.

Somehow, by careful adjustment of the quadrupoles in the lattice and by keeping the  $Q$ -spread (chromaticity) small, we must coax the beam up to full energy without hitting the lines. To make things more difficult, each line has a finite width, proportional to the strength of the imperfection which drives it. In some cases we must compensate the imperfections with correction multipoles to reduce this width.

But before discussing resonances and their correction in another contribution to these proceedings [4], a word about chromaticity.

## 8. CHROMATICITY

This steering of  $Q$  depends on careful regulation of quadrupole and dipole power supplies. In fact, much of the setting up time of a large accelerator is devoted to tune  $Q$  to be constant as the fields and energy rise. Once beam has been accelerated the problem becomes one of reducing all effects which produce a spread in  $Q$  among the particles in the beam. The limit to this is usually reached when beam intensity is high enough to cause space-charge focusing effects whose strength varies with the local beam density. Before reaching this limit one must correct the tune spread due to momentum: the chromaticity. This is exactly equivalent to the chromatic aberration in a lens. It is defined as a quantity  $\Delta Q'$

$$\Delta Q = Q' \frac{\Delta p}{p}.$$

It can be measured by changing the mean momentum of the beam by offsetting the r.f. frequency and measuring  $Q$ . Figure 22 shows such a measurement. Changing the r.f. frequency or momentum at a given field implies a change in radial position. As we have seen, an off-momentum particle will take up a new orbit following the dispersion function.

The chromaticity arises because the focusing strength of a quadrupole has  $(B\rho)$  in the denominator and is therefore inversely proportional to momentum. The focusing strength of the lattice quadrupoles,

$$k = \frac{1}{(B\rho)} \frac{dB_z}{dx} ,$$

varies inversely with  $(B\rho)$ , i.e., with momentum,  $p$ . A small spread in momentum in the beam,  $\pm\Delta p/p$ , causes a spread in focusing strength:

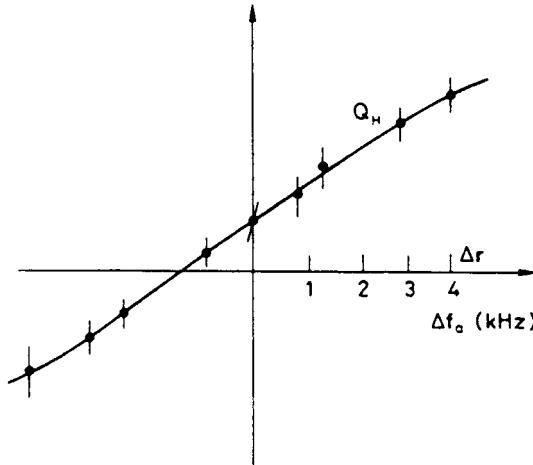


Fig. 22 Measurement of variation of  $Q$  with mean radius made by changing the r.f. frequency

$$\frac{\Delta k}{k} = -\frac{\Delta p}{p} .$$

Since the  $Q$ -value depends on  $k$ , we can also write a formula for the  $Q$ -spread:

$$\Delta Q = Q' \frac{\Delta p}{p} ,$$

where the constant  $Q'$  is the chromaticity, analogous to chromatic aberration in an optical system.

An equation we derived earlier in the section on gradient errors

$$\Delta Q = \frac{1}{4\pi} \int \beta(s) \delta k(s) ds .$$

enables us to calculate  $Q'$  rather quickly:

$$\Delta Q = \frac{1}{4\pi} \int \beta(s) \Delta k(s) ds = \left[ \frac{-1}{4\pi} \int \beta(s) k(s) ds \right] \frac{\Delta p}{p} .$$

The chromaticity  $Q'$  is just the quantity in square brackets. To be clear, this is called the natural chromaticity. For the SPS and the PS, indeed most AG machines, its value is about -1.3  $Q$  for H and V planes.

Imagine the situation at injection where  $\Delta p/p$  can be  $\pm 2 \times 10^{-3}$ . In a large synchrotron with a  $Q$  about 25 this can make the working point in the  $Q$  diagram into a line of length  $\Delta Q = 0.15$  which is too long to avoid the resonances. This must be corrected.

Just to make matters worse the chromaticity one has to correct may well be much greater than that due to the natural chromatic properties of quadrupoles. The remanent field at injection into a large ring may well be half a percent of the guide field and has the parabolic shape of a sextupole. In a superconducting ring the sextupole fields at injection stem from persistent currents and are very much larger still. Storage rings are usually designed with low-beta sections with zero dispersion for the interaction regions and the main low-beta quadrupoles being very strong make enormous contributions to the chromaticity. Since the dispersion is zero at the source of the error the compensation can only be made elsewhere in the lattice where the parameter  $D$  is large.

One way to correct this is to introduce some focusing which gets stronger for the high momentum orbits near the outside of the vacuum chamber – a quadrupole whose gradient increases with radial position is needed. Such a magnet has 6 poles, i.e. a sextupole. In a place where there is dispersion it will introduce a focusing

$$\Delta k = \frac{B'' D}{(B\rho)} \frac{\Delta p}{p} .$$

We use an earlier expression for the effect of this  $\Delta k$  on  $Q$  and obtain

$$\Delta Q = \left[ \frac{1}{4\pi} \int \frac{B''(s)\beta(s)D(s)ds}{(B\rho)} \right] \frac{dp}{p} .$$

To correct chromaticity we have merely to make the quantity in the square bracket balance the chromaticity.

However, there are two chromaticities, one affecting  $Q_x$ , the other  $Q_z$  and we must therefore arrange for the sextupoles to cancel both. For this we use a trick which is common in many different contexts. Sextupoles near F-quadrupoles where  $\beta_x$  is large affect mainly the horizontal  $Q$ , while those near D-quadrupoles where  $\beta_z$  is large influences  $Q_z$ . The effects of two families like this are not completely orthogonal but by inverting a simple 2 x 2 matrix one can find two sextupole sets which do the job.

The correction of chromaticity is a subject on its own since there is a higher-order term, a parabolic variation of  $Q$  with momentum which is not compensated in this way. Sextupole patterns which minimise this, yet do not themselves excite serious non-linear side effects, are not easy to find.

There are two ways of measuring chromaticity apart from the radial steering method shown in Fig. 22. The first of these is to observe the width of the betatron sidebands in the spectrum from a transverse pickup (Fig. 13). Secondly, we can measure the time it takes for a coherent betatron oscillation following a small kick to disappear as the  $\Delta Q$  smears out the phase relation between protons of different momenta (Fig. 23). A ringing time of 200 turns signifies a  $\Delta Q \approx 1/200$  and is about the best we can hope for using this rather crude method.

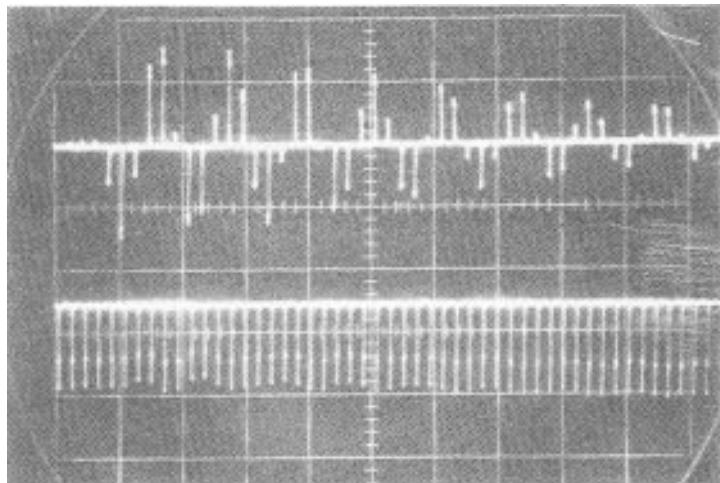


Fig. 23 Position pick-up signal following a kick showing decay of coherent betatron oscillation due to  $Q$  spread  $\approx 1/24$

## 9. CONCLUSIONS

We have now covered sufficient of the theory of transverse beam dynamics to understand the basic processes of designing focusing structures for a circular machine, the mechanisms which are produced by errors and how they may be compensated, at least in the case of those due to linear fields. The reader will also have learned enough about the behaviour of off-momentum particles to follow the explanation of longitudinal dynamics which follows. It remains only to list a bibliography of works which the author has found to be useful in understanding this topic.

\* \* \*

## REFERENCES

- [1] J. Rossbach and P. Schmüser, Basic course on accelerator optics, Proceedings of this School.
- [2] P.J. Bryant, Insertions, Proceedings of this School.
- [3] P.J. Bryant, Linear coupling, Proceedings of this School.
- [4] E. Wilson, Non-linearities and resonances, Proceedings of this School.
- [5] R.P. Walker, Synchrotron radiation, Proceedings of this School.
- [6] C.D. Johnson and L. Thorndahl, CERN/MPS/Int. Co. 68-13.
- [7] C. Bovet, R. Gouiran, I. Gumowski and H. Reich, A selection of formulae and data useful for the design of A.G. synchrotrons, CERN/MPS-SI/Int.DL/70/4.
- [8] B. Autin and P.J. Bryant, Closed orbit manipulation and correction in the CERN ISR, Proc. VIIIth Int. Conf. High Energy Acc. (CERN, Geneva, 1971).
- [9] E. Courant and H. Snyder, Theory of the alternating-gradient synchrotron, Ann. Phys. 3 (1958).
- [10] S. Guiducci, Chromaticity, Proceedings of this School.

## BIBLIOGRAPHY

- Livingood, J., 1961, Principles of Cyclic Accelerators (Van Nostrand).
- Livingston, M. and Blewett, J., 1962, Particle Accelerators (McGraw-Hill).
- Steffen, K.G., 1965, High Energy Beam Optics (Wiley).
- Bruck, H., 1968, Accelerateurs circulaire de particules (PUF, Paris).
- Rosenblatt, J., 1968, Particle Acceleration (Methuen).
- Kolomensky, A.A. and Lebedev, A.N., 1966, Theory of Cyclic Accelerators (North-Holland).
- Sands, M., 1970, The Physics of Electron Storage Rings, SLAC Report 121 (SLAC, Stanford, California).
- The 300 GeV Programme, 1972, CERN/1050 (CERN).
- Lawson, J.D., 1979, The Physics of Charged Particle Beams (Oxford).
- Theoretical Aspects of the Behaviour of Beams in Accelerators and Storage Rings, 1977, CERN 77-13 (CERN).
- Wilson, E.J.N., 1977, Proton Synchrotron Accelerator Theory, CERN 77-07 (CERN). FROM US
- Conte, M. and Mackay, W.W., An Introduction to the Physics of Particle Accelerators (World Scientific Publishing).
- Edwards, D.A. and Syphers, M.J., An Introduction to the Physics of High Energy Accelerators (Wiley).
- Bryant, P. and Johnsen, K., The Principles of Circular Accelerators and Storage Rings (Cambridge).

# INSERTIONS

*P.J. Bryant*  
CERN, Geneva, Switzerland

## Abstract

Following the advent of alternating-gradient (AG) focusing in 1952, there was a rapid transition from the monolithic  $360^0$  magnet of the old constant-gradient machines, to extended lattices using many individual magnetic elements. For a brief time, combined-function magnets were used, but separated-function lattices were soon to become the norm. AG focusing introduced such great flexibility that there has been a steady evolution in lattice design ever since leading to specialised lattices for different types of machines and dedicated insertions for specific tasks. This paper reviews insertions ranging from well-known examples, such as low- $\beta$  insertions, to lesser known examples, such as collimation insertions, and on to some more exotic examples, such as the monochromator.

## 1. INTRODUCTION

In the same year as the weak-focusing Cosmotron was finished at the Brookhaven National Laboratory (1952), Courant, Livingston and Snyder [1] proposed 'strong focusing', or 'alternating-gradient focusing', or more simply 'AG focusing'. The idea had been written down earlier by Christofilos [2], but it had not been published. This discovery marks the birth of the modern accelerator lattice. It revolutionised synchrotron design, allowing smaller magnets to be used and higher energies to be attained. It replaced the "single-unit" concept of the constant-gradient machine with its  $360^0$  monolithic magnet, by a "multi-segment" machine with a flexible lattice of individual magnetic elements. For a brief time, combined-function magnets were used, but separated-function lattices were soon to become the norm. Unlike many inventions, the introduction of AG focusing was both rapid and universal.

The principle of AG focusing is directly analogous to a well-known result in geometrical optics that a pair of spaced lenses of equal, but opposite strength, will be focusing overall. In fact, they will remain focusing over quite a large range of unequal strengths and spacings provided they remain with opposite signs. Intuitively, one is not surprised that this stability can be extended to longer series of alternating lenses and, furthermore, that the spacings and strengths can be greatly distorted to create special effects. This flexibility mixed with a great deal of ingenuity has led to a continuous development of specialised lattices for different machines and for specialised 'insertions' for use within a machine or a transfer line. In fact, a machine designer will now automatically think in modular terms of matched cells, achromats, dispersion suppressors, low- $\beta$  insertions and many other sub-units.

The design of a machine or transfer line starts with a basic lattice. This is usually a periodic series of matched cells that is long enough to incorporate all the required functions. The best known and most widely used matched cell is the FODO (F-focusing unit, D-defocusing unit and O-drift space or dipoles). The beam parameters within the cells are periodic and the lattice can be extended by simply adding cells (see Ref. [3] for an analysis). Some cells will be filled with dipoles while others will be left empty for rf cavities or for crossing long straight sections. To a first approximation the betatron focusing is left unaffected by the presence or absence of dipoles. Into this basic lattice at specific locations, the designer will 'insert' special functions that have become known as 'insertions'. The first is usually the dispersion suppressors that perform the transition between the periodic dispersion of the arcs to zero dispersion of the straight sections. Next he will add insertions for injection, extraction,

low- $\beta$  regions, small-angle crossings, etc. An insertion is designed to be matched into the basic lattice without perturbing the lattice functions outside itself. Finally, the designer will add correction and compensation schemes for the closed orbit, tune values, coupling and chromaticity. These schemes are distributed through the lattice and are not referred to as insertions.

The design of an insertion has two parts. Firstly it is necessary to define what special optical conditions are needed for a particular task. This may be evident in such cases as a low- $\beta$  insertion and less evident in such cases as a collimation insertion. Secondly, a match has to be made for the betatron and the dispersion functions from the special optical conditions to the basic lattice in which the insertion is embedded. The latter is a general problem and will be discussed first.

## 2. MATCHING

It is frequently necessary to match beam parameters over a short distance using quadrupole lenses, for example at each end of a transfer line or for a low- $\beta$  insertion in a collider. The problem is made difficult by the fact that while quadrupoles focus in one plane they defocus in the orthogonal plane. With just a few quadrupoles, it already becomes impossible to visualise how the betatron and dispersion functions will behave and 'back-of-envelope' calculations quickly become unmanageable. Some analytic solutions do exist and can be of considerable help in guiding computer computations. The early analytic work contains both thick- and thin-lens investigations [4, 5], but thin-lens solutions are more than adequate when backed-up by a computer program for numerical matching. The four examples presented here are a variable-geometry doublet, two triplets and a quadruplet, which use the lens positions as well as their strengths to obtain a match\*. This needs a stable optics environment, as in a collider, but if it is known that many widely different optics will be needed, as for example in a switch yard, a fixed-geometry solution must be adopted in which only the gradients are varied.

### 2.1 Variable-geometry doublet in a dispersion-free region [6]

Figure 1 shows the layout of a quadrupole doublet. Gradients and positions are taken as variables in order to match the well-known optical parameters  $\beta$  and  $\alpha$  in both planes. The insertion is assumed bending-free and dispersion-free ( $D = D' = 0$ ).

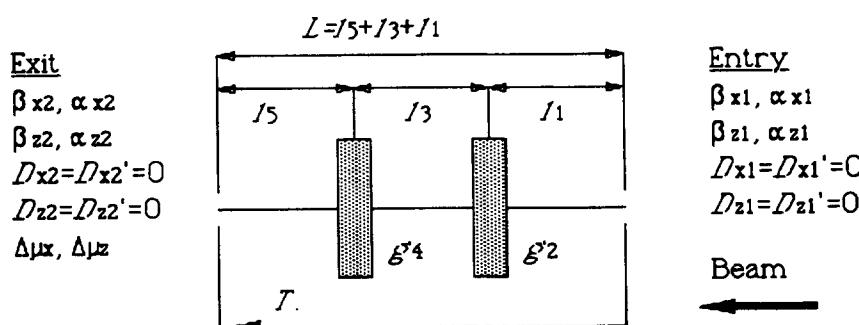


Fig. 1 Quadrupole doublet

\* The examples given here have been programmed as part of a package named AGILE230 (Alternating Gradient Interactive Lattice dEsign) for the IBM PC, or compatibles. The program is menu-driven and will perform many of the usual lattice calculations. The compiled and source versions are available from the author.

The doublet has only five free variables  $(\ell_1, g_2, \ell_3, g_4, \ell_5)$ , while the beam is described by eight matrix elements, which have six degrees of freedom (the condition for unit determinants reduces the degrees of freedom by two). Thus it is not possible to find a solution for all sets of beam parameters. Consistent solutions will only exist when the transfer matrix elements are related in a suitable way.

The elements of the transfer matrices of the doublet can be written in terms of the boundary conditions (for each plane) in the general form:

$$T(s_2 | s_1) = \begin{pmatrix} (\beta_2 / \beta_1)^{\frac{1}{2}}(\cos \Delta\mu + \alpha_1 \sin \Delta\mu) & (\beta_1 \beta_2)^{\frac{1}{2}} \sin \Delta\mu \\ -(\beta_1 \beta_2)^{-\frac{1}{2}}[(1 + \alpha_1 \alpha_2) \sin \Delta\mu + (\alpha_2 - \alpha_1) \cos \Delta\mu] & (\beta_1 / \beta_2)^{\frac{1}{2}}(\cos \Delta\mu - \alpha_2 \sin \Delta\mu) \end{pmatrix} \quad (1)$$

where  $\Delta\mu$  is the betatron phase shift through the doublet and is considered as a free parameter in both planes. The same transfer matrices can be expressed in terms of the thin lens expressions for the elements.

$$T_{h,v} = \begin{pmatrix} [1 \pm \ell_3 g_2 \pm \ell_5 g_2 \pm \ell_5 g_4 + \ell_3 \ell_5 g_2 g_4] & [\ell_1 + \ell_3 + \ell_5 \pm \ell_1 \ell_3 g_2 \pm \ell_1 \ell_5 g_2] \\ [\pm g_2 \pm g_4 + \ell_3 g_2 g_4] & [\pm \ell_1 \ell_5 g_4 \pm \ell_3 \ell_5 g_4 + \ell_1 \ell_3 \ell_5 g_2 g_4] \\ [\pm g_2 \pm g_4 + \ell_3 g_2 g_4] & [1 \pm \ell_1 g_2 \pm \ell_1 g_4 \pm \ell_3 g_4 + \ell_1 \ell_3 g_2 g_4] \end{pmatrix} \quad (2)$$

where the upper sign corresponds to the horizontal plane and the lower sign to the vertical. The inter-lens spaces are  $\ell_1, \ell_3$  and  $\ell_5$  and the integrated lens strengths are  $g_2$  and  $g_4$ . The comparison of the terms in (1) and (2) leads to the solution of the required parameters.

In order to have compact expressions, sum and difference matrices  $T^+, T^-$  are first defined whose elements are given by,

$$t_{i,j}^+ = \frac{1}{2}(t_{h,i,j} + t_{v,i,j}) \quad \text{and} \quad t_{i,j}^- = \frac{1}{2}(t_{h,i,j} - t_{v,i,j}) . \quad (3)$$

By means of (3), the horizontal and vertical transfer matrices (2) can be re-expressed as,

$$\begin{aligned} T^+ &= \begin{pmatrix} [1 + \ell_3 \ell_5 g_2 g_4] & [\ell_1 + \ell_3 + \ell_5 + \ell_1 \ell_3 \ell_5 g_2 g_4] \\ [\ell_3 g_2 g_4] & [1 + \ell_1 \ell_3 g_2 g_4] \end{pmatrix} \\ T^- &= \begin{pmatrix} [\ell_3 g_2 + \ell_5 g_2 + \ell_5 g_4] & [\ell_1 \ell_3 g_2 + \ell_1 \ell_5 g_2 + \ell_1 \ell_5 g_4 + \ell_3 \ell_5 g_4] \\ [g_2 + g_4] & [\ell_1 g_2 + \ell_1 g_4 + \ell_3 g_4] \end{pmatrix} \end{aligned} \quad (4)$$

and similarly the elements of  $T^+$  and  $T^-$  can be found in terms of the boundary conditions from (1). Direct inspection of matrices (4) leads to the solutions for the doublet parameters of,

$$\begin{aligned} \ell_1 &= (t_{22}^+ - 1) / t_{21}^+ & L &= (t_{11}^- + t_{22}^-) / t_{21}^- \\ \ell_5 &= (t_{11}^+ - 1) / t_{21}^+ & \ell_3 &= L - \ell_1 - \ell_5 . \end{aligned} \quad (5)$$

A little manipulation of elements  $t_{21}^-$  and  $t_{22}^+$  yields a quadratic equation for the gradients, from which the two solutions are found,

$$g_{2,4} = \frac{1}{2} [t_{21}^- \pm \sqrt{(t_{21}^-)^2 - 4t_{21}^+ / \ell_3}] \quad (g_2 \text{ takes } +\text{sign}) . \quad (6)$$

The insertion parameters expressed in (5) and (6) will be non-physical in certain cases. The distances must of course be positive and the gradients in (6) must be real, i.e.

$$(t_{21}^-)^2 \geq 4t_{21}^+ / \ell_3 . \quad (7)$$

However, this is still not sufficient to ensure a solution since there is one more equation to satisfy than parameters to vary. A further constraint must therefore be imposed, which is

$$(t_{21}^+)^2 t_{12}^- - (t_{22}^+ - 1)t_{11}^- t_{21}^+ - (t_{11}^+ - 1)t_{22}^- t_{21}^+ + (t_{22}^+ - 1)(t_{11}^+ - 1)t_{21}^- = 0 . \quad (8)$$

This condition is found by substitution into  $t_{12}^-$ , which has remained unused up to this point. It is assumed that the phase advances can be varied at will in (1) in order to satisfy (or nearly satisfy) this condition. The evaluation of (8) acts as a mismatch factor and as a guide when searching for a solution. The thin-lens result thus obtained can then be used in a standard lattice program as the starting point for optimising a 'thick' lens doublet.

## 2.2 First variable-geometry triplet

Figure 2 shows the layout of a simple triplet. Note that the elements have been numbered from  $g_2$  in order that the results can be used directly in the next section where drift spaces will be added on either side to increase the number of variable parameters.

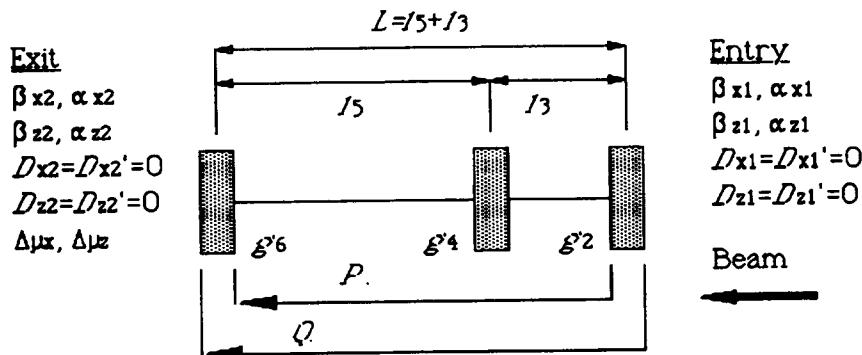


Fig. 2 First variable-geometry triplet

The same basic principles are applied to solving this triplet as for the doublet, except on this occasion it is easier to solve the parameter matrices in two stages. The central elements comprising the  $P$  matrix give,

$$P_{h,v} = \begin{pmatrix} 1 \pm g_4 \ell_5 & \ell_3 \pm g_4 \ell_3 \ell_5 + \ell_5 \\ \pm g_4 & 1 \pm g_4 \ell_3 \end{pmatrix}$$

which can be re-expressed in the sum-difference format as,

$$P^+ = \begin{pmatrix} 1 & \ell_3 + \ell_5 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad P^- = \begin{pmatrix} g_4 \ell_5 & g_4 \ell_3 \ell_5 \\ g_4 & g_4 \ell_3 \end{pmatrix} . \quad (9)$$

Equation (9) can be easily inverted to give,

$$g_4 = p_{21}^- ; \quad \ell_3 = p_{22}^- / p_{21}^- = p_{22}^- / g_4 ; \quad \ell_5 = p_{11}^- / p_{21}^- = p_{11}^- / g_4 \quad (10)$$

and

$$p_{11}^+ = p_{22}^+ = 1 ; p_{21}^+ = 0 ; \quad p_{12}^+ = \ell_3 + \ell_5 . \quad (11)$$

The overall matrices for the whole triplet are then given by,

$$\begin{aligned} Q^+ &= \begin{pmatrix} p_{12}^+ + g_2 p_{12}^- & p_{12}^+ \\ p_{21}^+ + g_6 p_{11}^- + g_2 p_{22}^- + g_2 g_6 p_{12}^+ & p_{22}^+ + g_6 p_{12}^- \end{pmatrix} \\ Q^- &= \begin{pmatrix} p_{11}^- + g_2 p_{12}^+ & p_{12}^- \\ p_{21}^- + g_6 p_{11}^+ + g_2 p_{22}^+ + g_2 g_6 p_{12}^- & p_{22}^- + g_6 p_{12}^+ \end{pmatrix}. \end{aligned} \quad (12)$$

In order to arrive at this result it should be noted that,

Main terms	Sum term	Diff. term
$p_{ij} \rightarrow$	$p_{ij}^+$	$p_{ij}^-$
$g_2 p_{ij} \rightarrow$	$g_2 p_{ij}^-$	$g_2 p_{ij}^+$
$g_2 g_6 p_{ij} \rightarrow$	$g_2 g_6 p_{ij}^+$	$g_2 g_6 p_{ij}^-$
$\ell_3 p_{ij} \rightarrow$	$\ell_3 p_{ij}^+$	$\ell_3 p_{ij}^-$

Equations (12) can be inverted, with the help of (11) to give,

$$\begin{aligned} g_2 &= \left( \frac{q_{11}^+ - p_{11}^+}{p_{12}^-} \right) = \left( \frac{q_{11}^+ - 1}{p_{12}^-} \right) = (q_{11}^+ - 1) / q_{12}^- \\ g_6 &= \left( \frac{q_{22}^+ - p_{22}^+}{p_{12}^-} \right) = \left( \frac{q_{22}^+ - 1}{p_{12}^-} \right) = (q_{22}^+ - 1) / q_{12}^- . \end{aligned} \quad (13)$$

With some more manipulation it is now possible to re-express the intermediate results (10) in terms of the overall matrix coefficients as,

$$g_4 = q_{21}^- - g_6 - g_2 - g_2 g_6 q_{12}^- ; \quad \ell_3 = (q_{22}^- - g_6 q_{12}^+) / g_4 \quad \text{and} \quad \ell_5 = (q_{11}^- - g_2 q_{12}^+) / g_4 . \quad (14)$$

As with the doublet there are insufficient parameters to unambiguously satisfy all boundary conditions and an extra constraint is needed. This is found by substitution into the  $q_{21}^+$  coefficient, which has, as yet, not been used.

$$\begin{aligned} &(q_{22}^+ - 1)(q_{11}^- q_{12}^- - q_{12}^+ q_{11}^+ + q_{12}^+) + (q_{11}^+ - 1)(q_{22}^- q_{12}^- - q_{12}^+ q_{22}^+ + q_{12}^+) \\ &+ (q_{11}^+ - 1)(q_{22}^+ - 1)q_{12}^+ - (q_{12}^-)^2 q_{21}^+ = 0 . \end{aligned} \quad (15)$$

Equation (15) can be treated as a mismatch factor when searching for a true solution. This is done in the same way as for the doublet in the previous section by varying the horizontal and vertical phase advances until the mismatch factor (15) is zero and suitable parameters for the triplet have been obtained.

### 2.3 Second variable-geometry triplet

Figure 3 shows the layout of a triplet with two more variable parameters. The extra flexibility makes it possible to find exact solutions directly and to also match the dispersion. An analytic solution to this triplet can be found in Ref. [7].

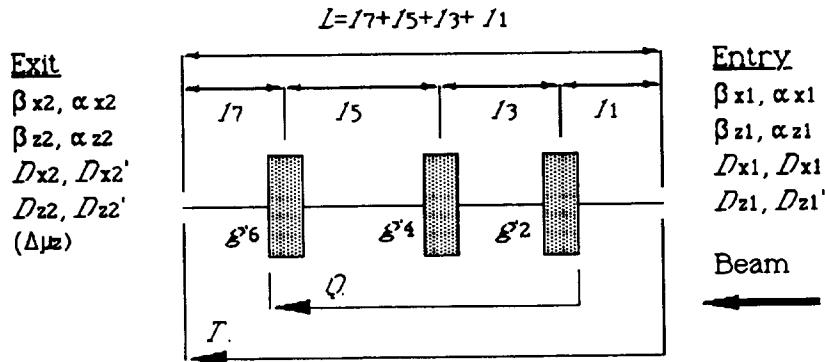


Fig. 3 Second variable-geometry triplet

The addition of two drift spaces appears simple, but it vastly complicates the algebra. With the inclusion of the outer drift spaces the new sum and difference matrices  $T^+$  and  $T^-$  for the overall insertion are given in (16). The  $Q^+$  and  $Q^-$  matrices can be taken directly from the previous section.

$$T^\pm = \begin{pmatrix} q_{11}^\pm + \ell_7 q_{21}^\pm & \ell_1 q_{11}^\pm + q_{12}^\pm + \ell_1 \ell_7 q_{21}^\pm + \ell_7 q_{22}^\pm \\ q_{21}^\pm & \ell_1 q_{21}^\pm + q_{22}^\pm \end{pmatrix} \quad (16)$$

The  $T^+$  and  $T^-$  matrices combined in (16) can be inverted to give,

$$\begin{aligned} q_{21}^\pm &= t_{21}^\pm \\ q_{11}^\pm &= t_{11}^\pm - \ell_7 q_{21}^\pm = t_{11}^\pm - \ell_7 t_{21}^\pm \\ q_{22}^\pm &= t_{22}^\pm - \ell_1 q_{21}^\pm = t_{22}^\pm - \ell_1 t_{21}^\pm \\ q_{12}^\pm &= t_{12}^\pm - \ell_1 t_{11}^\pm - \ell_7 t_{22}^\pm - \ell_1 \ell_7 t_{21}^\pm \end{aligned} \quad (17)$$

Thus all the  $q_{ij}$ 's are known if  $\ell_1$  and  $\ell_7$  are fixed. It is convenient to redefine  $\ell_1$  in terms of the total length  $L$ ,

$$\ell_1 = L - \ell_3 - \ell_5 - \ell_7 . \quad (18)$$

A further relation is now needed to establish  $\ell_7$ . This is achieved by substitution into the boundary constraint (15) in the previous section. The algebra is very tedious but results in a quadratic that determines the values of  $\ell_7$  that will ensure the boundary condition is true.

$$\begin{aligned} &\ell_7^2 (ebt_{21}^- - edt_{21}^+) + \ell_7 (abt_{21}^+ + eat_{21}^- - ebt_{11}^- - ect_{21}^+ + def + fb^2) \\ &+ (a^2 t_{21}^+ - eat_{11}^- + ecf + fab) = 0 \end{aligned} \quad (19)$$

where

$$\begin{aligned} a &= t_{12}^- - \ell_1 t_{11}^- ; \quad b = \ell_1 t_{21}^- - t_{22}^- ; \quad c = t_{12}^+ - \ell_1 t_{11}^+ \\ d &= \ell_1 t_{21}^+ - t_{22}^+ ; \quad e = t_{22}^+ - \ell_1 t_{21}^+ - 1 ; \quad f = t_{11}^+ - 1 . \end{aligned} \quad (20)$$

Now that  $\ell_7$  is known all the elements of the  $Q$  matrices can be found and by substitution into the results obtained in the previous sections the insertion parameter can be determined.

$$\begin{aligned} g_2 &= (f - \ell_7 t_{21}^+) / (a + \ell_7 b) & , \quad g_6 &= e / (a + \ell_7 b) \\ g_4 &= t_{21}^- - g_2 - g_6 - g_2 g_6 (a + \ell_7 b) & , \\ \ell_3 &= [-b - g_6 (c + \ell_7 d)] / g_4 & , \quad \ell_5 &= [t_{11}^- - \ell_7 t_{21}^- - g_2 (c + \ell_7 d)] / g_4 . \end{aligned} \quad (21)$$

A search is then made for suitable solutions by introducing a range of values for  $\ell_1$  and  $\Delta\mu_z$ . Since there are now seven variables, it is possible to satisfy the boundary condition (15) and all solutions will be exact. Horizontal dispersion matching is introduced by the restriction this imposes on the horizontal phase advance through the insertion, which is given by,

$$\begin{aligned} \cos \Delta\mu &= \frac{D_1 D_2 + (\alpha_1 D_1 + \beta_1 D'_1)(\alpha_2 D_2 + \beta_2 D'_2)}{(\beta_1 \beta_2)^{1/2} H} \\ \sin \Delta\mu &= \frac{D_2 (\alpha_1 D_1 + \beta_1 D'_1) - D_1 (\alpha_2 D_2 + \beta_2 D'_2)}{(\beta_1 \beta_2)^{1/2} H} . \end{aligned} \quad (22)$$

Equations (22) can be derived by considering the dispersion function in a bending-free region as a betatron oscillation. The function  $H$  is defined below as,

$$H = \gamma D(s)^2 + 2\alpha D(s)D'(s) + \beta D'(s)^2 \quad (= \text{constant}) . \quad (23)$$

It should be noted that  $H$  is an invariant (dispersion invariant) in a bending-free region. Equations (22) and (23) are sometimes of use when designing an injection insertion and transforming beam parameters from septum to kicker.

## 2.4 Variable-geometry quadruplet

Figure 4 shows the layout of a simple quadruplet. This insertion has the same number of variables as the previous triplet and can therefore also be used to match dispersion.

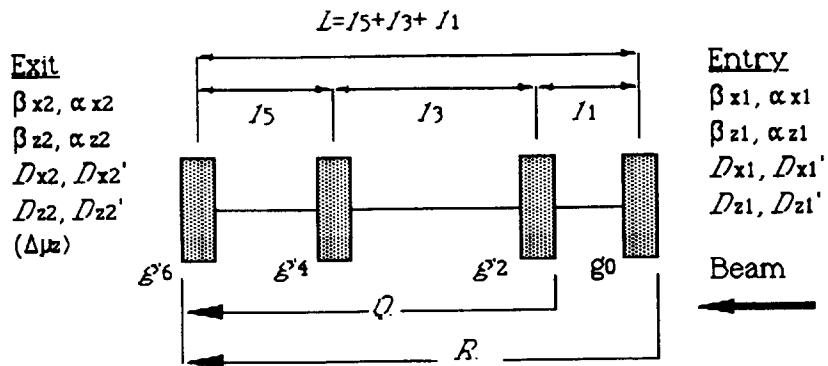


Fig. 4 Simple variable-geometry quadruplet

In order to analyse this insertion recourse is again made to the results already obtained for the simple variable-geometry triplet. The overall transfer matrices are then found by multiplying in the extra quadrupole and drift space at the entry to give,

$$R^\pm = \begin{pmatrix} q_{11}^\pm + g_0 \ell_1 q_{11}^\mp + g_0 q_{12}^\mp & \ell_1 q_{11}^\pm + q_{12}^\pm \\ q_{21}^\pm + g_0 \ell_1 q_{21}^\mp + g_0 q_{22}^\mp & \ell_1 q_{21}^\pm + q_{22}^\pm \end{pmatrix}. \quad (24)$$

The inversion of (24) enables all the elements of  $Q$  to be expressed in terms of  $g_0$ ,  $\ell_1$  and the elements of  $R$ ,

$$\begin{aligned} q_{11}^+ &= r_{11}^+ - g_0 r_{12}^- = c \\ q_{11}^- &= r_{11}^- - g_0 r_{12}^+ = d \\ q_{21}^+ &= r_{21}^+ - g_0 r_{22}^- = a \\ q_{21}^- &= r_{21}^- - g_0 r_{22}^+ = b \\ q_{22}^+ &= r_{22}^+ - \ell_1(r_{21}^+ - g_0 r_{22}^-) = r_{22}^+ - \ell_1 a \\ q_{22}^- &= r_{22}^- - \ell_1(r_{21}^- - g_0 r_{22}^+) = r_{22}^- - \ell_1 b \\ q_{12}^+ &= r_{12}^+ - \ell_1(r_{11}^+ - g_0 r_{12}^+) = r_{12}^+ - \ell_1 c \\ q_{12}^- &= r_{12}^- - \ell_1(r_{11}^- - g_0 r_{12}^+) = r_{12}^- - \ell_1 d. \end{aligned} \quad (25)$$

The strategy is to choose a value for  $g_0$  and to solve for  $\ell_1$  by substitution into the boundary constraint (15). This is a very long and tedious algebraic task, but it leads to a quadratic in  $\ell_1$  given in (26). Thus  $\ell_1$  is fixed in such a way that the boundary constraint (15) is satisfied, so all solutions will be exact. If the phase advances are now fixed, either by the dispersion when it is non-zero or by choice, the quadruplet can be solved.

$$\begin{aligned} \ell_1^2(fbd + aj - ad^2) + \ell_1(2adr_{12}^- - ah - eg - fbr_{12}^- - fdr_{22}^-) \\ + (eh + fr_{22}^- r_{12}^- - (r_{12}^-)^2 a) = 0. \end{aligned} \quad (26)$$

Once the values of  $\ell_1$  are established by solving the above quadratic, the remaining parameters can be found by back substitution into the earlier expressions to give,

$$\begin{aligned} g_2 &= f / (r_{12}^- - \ell_1 d) \\ g_6 &= (e - \ell_1 a) / (r_{12}^- - \ell_1 d) \\ g_4 &= b - g_2 - g_6 - g_2 g_6 (r_{12}^- - \ell_1 d) \\ \ell_3 &= [(r_{22}^- - \ell_1 b) - g_6 (r_{12}^+ - \ell_1 c)] / g_4 \\ \ell_5 &= [d - g_2 (r_{12}^+ - \ell_1 c)] / g_4 \end{aligned} \quad (27)$$

where the addition symbols are defined as,

$$\begin{aligned} e &= r_{22}^+ - 1 & f &= c - 1 \\ h &= dr_{12}^- - fr_{12}^+ & j &= d^2 - c^2 + c. \end{aligned} \quad (28)$$

By treating  $g_0$  and one, or both, of the phase advances as free parameters a range of solutions can be calculated and the most suitable selected.

## 2.5 Numerical matching

Matching by numerical minimisation will always be needed. Very often, it will be the only tool available for solving a problem, but even when an analytic solution exists it will still need numerical optimisation. Powerful lattice programs such as those in Refs. 8 to 14 have therefore been developed. Most numerical matching programs will move elements, alter their strength, length and edge-focusing angles, while minimising mismatches, maximum values and perhaps also imposing user specified conditions such as ratios between optical parameters. They will do this by varying several lattice parameters, which can be single elements or strings or specified combinations of elements. The starting point for the lattice parameters is important and it is here that some analytic indication of a solution helps the user to avoid being trapped in a local minimum or chasing after a condition that is fundamentally impossible.

The efficiency of numerical matching depends to some degree on the experience of the user and his familiarity with the program. When difficulties are met, it is necessary to be as inventive as possible to coax the program to look in regions of the parameter space that it has so far ignored. The first step is to vary the weights on the parameters being matched. Dispersion matching is often more difficult than betatron matching, so it can be advantageous to apply far more weight to the dispersion first and then to slowly shift to the betatron matching once the dispersion is matched. It is sometimes possible to 'walk' from one solution, which is say good except for one parameter, by making very small steps in that parameter and rematching each time. Trial and error changes in the starting point may help the program to find a new minimum. In principle, an equal number of variables and constraints are needed, but in general variable gradients will be found more useful than variable positions. Unfortunately, if no solution is found it is not sure that one does not exist and if one is found it is not sure it is the only one nor that it is the best. Numerical matching programs have the very useful flexibility of being able to distort existing solutions, for example, to see if two magnets can be put in series, or whether a magnet can be moved to make room for a monitor and so on.

## 3. DISPERSION SUPPRESSORS EMBEDDED IN A FODO LATTICE [15]

It has already been mentioned that the presence of the dipoles in a FODO cell makes little difference to its betatron focusing properties. The dispersion function, however, will sense the missing dipoles and become mismatched. The most convenient solution therefore would be to suppress the dispersion at the exit to an arc, so that it remains zero until the next bend. This can be solved in a rather elegant way by acting only on the dipoles in the FODO cells.

The Fig. 5 shows a general layout of a dispersion suppressor. The regular arc is on the left, the straight section on the right, and in between there is the dispersion suppressor insertion of  $N$  cells. This comprises a 'missing dipole' gap of  $(N-i)$  cells and a section of  $i$  cells with a special bending strength. In some case, the 'missing dipole' gap will be of zero length.

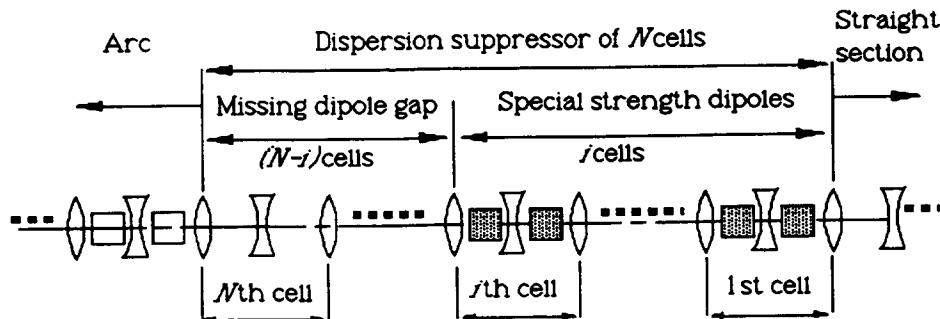


Fig. 5 General layout of a dispersion suppressor  
(Note 'missing dipole' gap may be of zero length)

A solution for the dispersion function is needed that starts with  $D = D' = 0$  at the exit to the straight section and builds up over the  $N$  FODO cells of the suppressor to the matched periodic value for the dispersion at the entry to the arc. The starting point is a general formulation of the dispersion function that can be found in Ref. [3].

$$\begin{aligned} D(2NL) &= S(2NL) \int_0^{2NL} \frac{C(s)}{\rho(s)} ds - C(2NL) \int_0^{2NL} \frac{S(s)}{\rho(s)} ds \\ D'(2NL) &= S'(2NL) \int_0^{2NL} \frac{C(s)}{\rho(s)} ds - C'(2NL) \int_0^{2NL} \frac{S(s)}{\rho(s)} ds \end{aligned} \quad (29)$$

where  $L$  is the half length of a FODO cell and  $\rho(s)$  is the bending radius in the various cells. The functions  $C(s)$ ,  $S(s)$ ,  $C'(s)$  and  $S'(s)$  are known as the principal trajectories and are also explained in Ref. [3].

The limits of the insertion are taken at the centres of the lenses. This can either be mid-D to mid-D  $N$  cells later or mid-F to mid-F, but in either case the transfer matrix for the insertion can be written as,

$$T_{N \text{ cells}} = \begin{pmatrix} \cos(N\mu_0) & \tilde{\beta} \sin(N\mu_0) \\ -\tilde{\beta}^{-1} \sin(N\mu_0) & \cos(N\mu_0) \end{pmatrix} = \begin{pmatrix} C(2NL) & S(2NL) \\ C'(2NL) & S'(2NL) \end{pmatrix} \quad (30)$$

where  $\tilde{\beta}^{-1}$  represents  $\tilde{\beta}$  when the insertion is mid-F to mid-F and  $\check{\beta}$  when the insertion is mid-D to mid-D. The phase advance  $\mu_0$  corresponds to a full FODO cell. Suffixes to indicate the plane are omitted. The theory applies equally to both planes but for all machines will be in the horizontal plane. Equation (30) gives the multipliers  $S(2NL)$  etc. of the integrals in (29).

The evaluation of the integrals in (29) appears complicated, but it is sufficient to use mid-point values for  $C(s)$  and  $S(s)$  in each half cell. These can be found by applying the general transfer matrix through a lattice, which is also to be found in Ref. [3].

$$T(\text{mid-pt. of dipole} \mid s = 0) = \begin{pmatrix} (\beta_m / \tilde{\beta})^{1/2} \cos \mu_n & (\beta_m \tilde{\beta})^{1/2} \sin \mu_n \\ (\text{not needed}) & (\text{not needed}) \end{pmatrix} \quad (31)$$

where  $\beta_m$  is the betatron amplitude at the mid-point of the dipole in each half cell and  $\mu_n$  is the phase up to each dipole mid-point from the entry to the dispersion suppressor. This can be expressed as,

$$\mu_n = (n - 1/2)\mu_0 \pm \Delta\mu \quad (32)$$

where  $\Delta\mu$  is the phase shift from the centre quadrupole of a cell upstream or downstream to the adjacent mid-point of the half-cell dipole.

From the above the integrals (29) can be rewritten as sums over the mid-points of the dipoles in the  $i$  filled cells, where  $i$  in some cases will assume the maximum value of  $N$ .

$$\begin{aligned}
D(2NL) &= \left(\tilde{\beta}\beta_m\right)^{1/2} \left[ \sin(N\mu_0) \sum_{n=1}^i \left\{ \delta_{n-} \cos[(n-1/2)\mu_0 - \Delta\mu] + \delta_{n+} \cos[(n-1/2)\mu_0 + \Delta\mu] \right\} \right. \\
&\quad \left. - \cos(N\mu_0) \sum_{n=1}^i \left\{ \delta_{n-} \sin[(n-1/2)\mu_0 - \Delta\mu] + \delta_{n+} \sin[(n-1/2)\mu_0 + \Delta\mu] \right\} \right] \\
D'(2NL) &= \left(\frac{\beta_m}{\tilde{\beta}}\right)^{1/2} \left[ \cos(N\mu_0) \sum_{n=1}^i \left\{ \delta_{n-} \cos[(n-1/2)\mu_0 - \Delta\mu] + \delta_{n+} \cos[(n-1/2)\mu_0 + \Delta\mu] \right\} \right. \\
&\quad \left. + \sin(N\mu_0) \sum_{n=1}^i \left\{ \delta_{n-} \sin[(n-1/2)\mu_0 - \Delta\mu] + \delta_{n+} \sin[(n-1/2)\mu_0 + \Delta\mu] \right\} \right]
\end{aligned} \tag{33}$$

where  $\delta_{n-}$  and  $\delta_{n+}$  are the angular kicks ( $\delta = L/\rho$ ) in the dipoles of the dispersion suppressor upstream and downstream of the centre quadrupole of the  $n$ th FODO cell. Equations (33) simplify to give,

$$\begin{aligned}
D(2NL) &= \left(\tilde{\beta}\beta_m\right)^{1/2} \left[ \sum_{n=1}^i \left\{ \delta_{n-} \sin[(N-n+1/2)\mu_0 + \Delta\mu] + \delta_{n+} \sin[(N-n+1/2)\mu_0 - \Delta\mu] \right\} \right] \\
D'(2NL) &= \left(\frac{\beta_m}{\tilde{\beta}}\right)^{1/2} \left[ \sum_{n=1}^i \left\{ \delta_{n-} \cos[(N-n+1/2)\mu_0 + \Delta\mu] + \delta_{n+} \cos[(N-n+1/2)\mu_0 - \Delta\mu] \right\} \right].
\end{aligned} \tag{34}$$

In order to set both  $D$  and  $D'$  to the desired values, it will be assumed that all dipoles in the suppressor have the same strength and this will be taken as the first free variable. The values of  $i$  and  $N$  will then be taken as additional variables. Once the strengths of the dipoles are taken as equal, (34) can be further simplified to,

$$\begin{aligned}
D(2NL) &= 2\left(\tilde{\beta}\beta_m\right)^{1/2} \delta \cos(\Delta\mu) \sum_{n=1}^i \sin[(N-n+1/2)\mu_0] \\
D'(2NL) &= 2\left(\frac{\beta_m}{\tilde{\beta}}\right)^{1/2} \delta \cos(\Delta\mu) \sum_{n=1}^i \cos[(N-n+1/2)\mu_0].
\end{aligned} \tag{35}$$

The standard trigonometrical series in Ref. [16] can be used to sum equations (35) to give,

$$D(2NL) = 2\left(\tilde{\beta}\beta_m\right)^{1/2} \delta \cos(\Delta\mu) \frac{\sin[(N-i/2)\mu_0] \sin(i\mu_0/2)}{\sin(\mu_0/2)} \tag{36}$$

$$\text{and } D'(2NL) = 2\left(\frac{\beta_m}{\tilde{\beta}}\right)^{1/2} \delta \cos(\Delta\mu) \frac{\cos[(N-i/2)\mu_0] \sin(i\mu_0/2)}{\sin(\mu_0/2)} \tag{37}$$

where  $0 < i \leq N$ . Equation (36) must now be equated to the matched value of the dispersion in the arc. If the insertion is between mid-F points then this will be the maximum dispersion in the arc and if it is between mid-D points it will be the minimum. In both cases, (37) should be equated to zero, which requires either  $\sin(i\mu_0/2) = 0$  or  $\cos[(N-i/2)\mu_0] = 0$ . The second condition is the more useful and yields,

$$\mu_0(N-i/2) = (2j-1)\pi/2 \quad \text{where } j = 1, 2, 3, \dots \tag{38}$$

Equation (36) is now used to evaluate the strength  $\delta$  of the dipole kicks. For this, it is necessary to find  $\cos(\Delta\mu)$  the phase shift from the mid-point of the central quadrupole to the mid-point of the adjacent dipole. This can be done by using the  $t_{12}$  term in matrix (31) and equating it to the same term evaluated using the sub-matrices for a thin lens and a drift space of length  $L/2$  to represent the dipole. The thin lens approximation is described in Ref. [3].

$$(\beta_m \tilde{\beta})^{1/2} \sin(\Delta\mu) = L/2 \quad (39)$$

so that  $\cos(\Delta\mu) = \sqrt{1 - L^2 / (4\beta\tilde{\beta})}$ . (40)

Equation (36) can now be evaluated using the well-known expressions, reproduced below, for the maximum and minimum values of  $D$  and  $\beta$  and the mid-point  $\beta_m$  in a FODO cell. For further details see Ref. [3])

$$\begin{aligned} \hat{\beta} &= 2L \frac{1 + \sin(\mu_0 / 2)}{\sin \mu_0} ; & \check{\beta} &= 2L \frac{1 - \sin(\mu_0 / 2)}{\sin \mu_0} ; \\ \beta_m &= \frac{2L}{\sin \mu_0} \left[ 1 - \frac{1}{2} \sin^2(\mu_0 / 2) \right] ; & & \\ \hat{D} &= \frac{(2L)^2}{\rho_0} \frac{\left[ 1 + \frac{1}{2} \sin(\mu_0 / 2) \right]}{4 \sin^2(\mu_0 / 2)} \quad \text{and} \quad \check{D} &= \frac{(2L)^2}{\rho_0} \frac{\left[ 1 - \frac{1}{2} \sin(\mu_0 / 2) \right]}{4 \sin^2(\mu_0 / 2)} . \end{aligned} \quad (41)$$

After some manipulation, the same result [see (42)] is obtained whether the insertion is made from mid-D to mid-D and matched to the minimum value of dispersion in the FODO cell of the arc, or if it is made mid-F to mid-F and matched to the maximum dispersion.

$$\frac{1}{2} \frac{L}{\rho_0} = \delta (-1)^{j+1} \sin(i \mu_0 / 2) \quad (42)$$

where  $\delta$  is the angular kick of each dipole in the  $i$  cells of the suppressor and  $L/\rho_0$  is the angular kick of each normal dipole in the arc. Both  $\delta$  and  $L/\rho_0$  correspond to a single dipole in a half FODO cell which has the whole half-cell length.

It is quickly verified that the two conditions set by (38) and (42) are satisfied by the examples given in Table 1, which are known as missing-magnet dispersion suppressors. The most convenient configuration is that in which the remaining dipoles retain their full value.

**Table 1**  
Examples of missing-magnet dispersion suppressors

N	i	n	$\mu_0$	$\delta$
2	1	1	60°	$(L/\rho_0)$
3	2	1	45°	$(L/\mu_0)/\sqrt{2}$
4	1	2	77°	$-(L/\rho_0)/1.25$
...	...	...	...	...

There is also the class of solutions which occur when  $N = i$ . This corresponds to there being no empty magnet gap and these suppressors rely only on having  $\delta$  different to the normal value in the arcs. In this case (38) reduces to,

$$N \mu_0 = (2j - 1)\pi \quad \text{where } j = 1, 2, 3, \dots \quad (43)$$

and (42), from which the dipole strength is derived, becomes,

$$\delta = \frac{1}{2} \frac{L}{\rho_0}. \quad (44)$$

This extremely simple result has given rise to the name reduced-field dispersion suppressors. Table 2 lists some specific solutions.

**Table 2**  
Examples of half-field dispersion suppressors

N	n	$\mu_0$	$\delta$
2	1	90°	$(L/\mu_0)/2$
3	1	60°	"
4	1	45°	"
5	2	108°	"
...	...	...	...

The simplicity of these solutions is at first sight remarkable. A simple gap with all dipoles of the standard strength, or a region of half-field dipoles are extremely practical solutions in terms of power supplies and magnet design. The half-field case is especially attractive since by leaving out half of the dipoles in a half cell, the remaining units can be moved longitudinally so as to give a fine adjustment to the matching. In all cases the dispersion stays well-behaved and reduces to zero in a regular manner so that there are no special aperture requirements within the insertions.

There are simple explanations for these solutions. In the 'missing magnet' case, the phase condition (38) has the physical meaning of requiring the phase shift from the mid-point of the group of i cells to the end of the suppressor to be an odd number of quarter betatron wavelengths. This scheme will therefore operate by kicking the beam and arranging for the subsequent oscillation of the dispersion function to reach its maximum at the end of the suppressor and to be equal to the matched value in the arc cells.

In the half-field suppressors, the dispersion coming from the arc finds itself mismatched in the half-field region and having twice its local matched value. Consequently it oscillates and half a wavelength later it has dropped to zero. By moving directly into the straight section the dispersion is matched to zero.

#### 4. LOW- $\beta$ INSERTIONS

The goal of a collider is to have a high rate of particle collisions and a small, stored-beam energy. The collision rate is described more precisely in terms of the 'luminosity', which is defined as 'the ratio of the event rate of a chosen interaction at a beam crossing to the cross-section of that interaction'. One way of achieving a high luminosity with a low-stored energy is to make the sizes of the two beams very small at the crossing point. For head-on collisions the beams must be small in both planes, but when there is a crossing angle only the beam height perpendicular to the plane of crossing need be reduced. The usefulness of the so-called 'low- $\beta$  insertion' for achieving this result was recognised many years ago [17]. Certain basic characteristics of the low- $\beta$  insertion are easily understood. The point of minimum beam cross-section is placed exactly at the crossing point. It is shown in [3] that the betatron amplitude will grow with the distance  $s$  from this point in the drift spaces on either side according to,

$$\beta(s) = \beta(0)[1 + s^2 / \beta^2(0)]. \quad (45)$$

The smaller  $\beta(0)$  is, the faster  $\beta(s)$  will grow with distance from the crossing point. This fundamental dependence is a limiting factor on the space that can be made available for physics detectors. One way of partially overcoming this problem is to mount the first quadrupole of the low- $\beta$  insertion inside the detector itself. By using an iron-free, superconducting magnet, the interference with the detector can be minimised and the gradient can still be strong. At the first quadrupole, one plane will be focused while the other will be defocused. The betatron amplitude in the defocused plane will therefore be increased even faster than before. For example, the CERN LHC (Large Hadron Collider) is designed to have a minimum betatron amplitude of 0.3m with a maximum in the second quadrupole of the insertion of nearly 5000m. Clearly the second quadrupole cannot be placed too far away from the first. This gives the characteristic layout of such an insertion with a closely grouped doublet or triplet set symmetrically either side of the crossing point (see Fig. 6). Apart from the physical problem of the increased beam size in the quadrupoles, such large betatron amplitudes are a source of extremely strong nonlinear effects, which arise because of the momentum spread in the beam. Since the shape of the betatron function is known in a drift space (45), it is easy to calculate the phase advance between the innermost quadrupoles. Thus from (45)

$$\Delta\mu_{\text{insertion}} = \frac{1}{\beta(0)} \int_{-L}^L \frac{ds}{1 + s^2 / \beta^2(0)} = 2 \tan^{-1}[L / \beta(0)]. \quad (46)$$

Usually  $L \gg \beta(0)$  so that the phase advance approaches  $\pi$ . Thus, the tune of the machine will increase by approximately a half-integer for each low- $\beta$  insertion added.

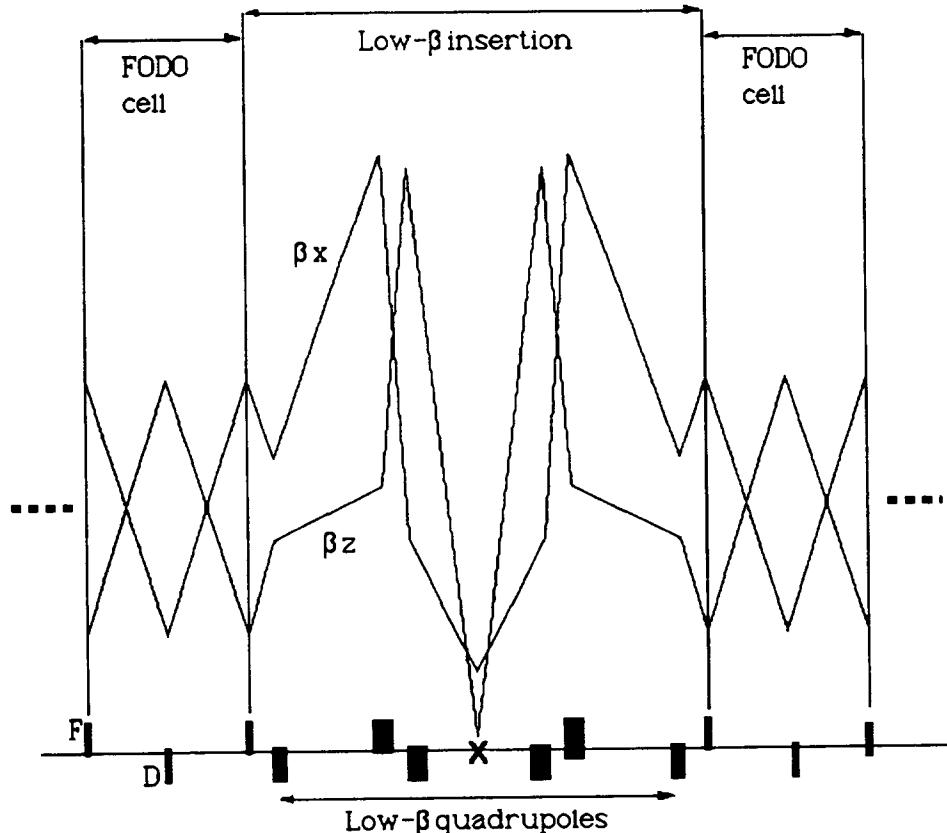


Fig. 6 Symmetric low- $\beta$  insertion matched to a FODO lattice (schematic)

Low- $\beta$  insertions with doublets, triplets or quadruplets can be calculated analytically using the thin-lens approximation described in Section 2. In general, a low- $\beta$  insertion will be in a dispersion-free region and a doublet may be sufficient to match to the main lattice, but the

triplet is more flexible and is usually used. When the dispersion is non-zero, a triplet can still be used, but it is likely that a quadruplet will be needed. In all cases, the final design must be done with a numerical matching program.

## 5. EXTRACTION AND INJECTION [18]

The most straightforward way of extracting, or injecting, a beam from a circular machine is to use a fast kicker magnet and a septum (see Fig. 7). The septum is a dipole magnet with a thin partition (septum plate) that isolates the septum's field from the main aperture of the circulating beam. When extracting a beam, the kicker is used to deflect the beam onto an orbit that oscillates away from the central orbit and into the aperture of the septum. When injecting a beam, the reverse is true and the kicker is used to deflect the incoming particle from its septum-kicker trajectory onto the central orbit. It is reasonably easy to track from the kicker to the septum and to match the amplitudes and angles of the orbits to the ring on one side and to the transfer line on the other. The quadrupoles in the transfer line are then used to match the beam ellipses. By inspection one can see that D-quadrupoles between the kicker and septum will work in favour of the scheme, it is also clear that a phase advance of about  $\pi/2$  will convert the angular kick of the kicker to a maximum amplitude at the septum. A simple solution would be to use a  $\pi/2$  FODO cell as illustrated in Fig. 7. The D-unit 'sees' a large amplitude and aids the deflection of the beam and the F-unit close to the kicker 'sees' such a small amplitude that it has little adverse effect. The problems are usually of a more technological nature. The extracted/injected beam will pass farther from the axis than normal and the lattice units may need larger apertures with detailed magnetic field measurements. Between the septum and the transfer line the beam may have to pass through stray fields or even through holes in magnet yokes and supports. The clean, full-aperture extraction of a beam during one turn of the machine requires a gap in the beam for the kicker rise time. If the normal bunch spacing is not enough, there may have to be missing bunches that will cause beam loading problems in the rf cavities. If single bunches are to be extracted or injected, then both the rise and fall times of the kicker are critical. If the kicker has insufficient strength, as often occurs, it may be possible to use a half-wave, closed-orbit bump to bring the beam closer to the septum and hence to reduce the required kick. Another possibility is to use an electrostatic septum. Electrostatic septa are weaker, but much thinner, so that the kicker needs less strength to cross the septum wires into the extraction aperture.

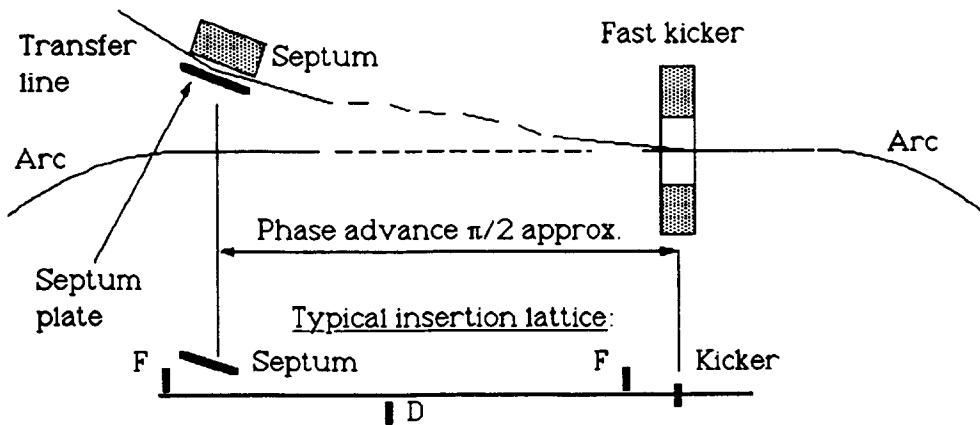


Fig. 7 Fast kicker and septum for an extraction or injection insertion

The simple use of a kicker and septum to inject bunches onto the central orbit does not always use the available phase space to the best advantage. Often higher currents can be stored by multi-turn schemes that inject bunches into different regions in the betatron phase space or the momentum phase space. The suitable phasing of a coherent oscillation at injection or a shrinking bump can be used for betatron stacking [19] and rf stacking can be used to fill the

momentum space as suggested by MURA [20] (Midwestern Universities Research Association) and successfully used at the CERN ISR (Intersecting Storage Rings) for many years.

A very different technique, that of charge 'exchange injection', was pioneered in Novosibirsk [21] and has become the preferred method for injection into high intensity machines. The first step is to produce a negative ion of the particle to be accelerated. The H<sup>-</sup> source is well known for injection into proton machines, but negative ions can be made of most elements. The negative-ion beam is injected and put onto the same orbit as the circulating beam (see Fig. 8). This is easily achieved since a dipole will bend the oppositely charged particles in opposite directions. In the drift space where the two beams coincide both beams pass through a thin foil, which strips away the weakly bound electrons from the negative ions to create positive particles for the main beam (protons in the case of an H<sup>-</sup> beam). Those particles that are either not stripped, or are stripped to the wrong charge state, will be separated from the circulating beam by the next dipole and can be trapped on an absorber block. In this way, the injected particles are placed in precisely the same phase space as the already circulating beam and by repeating the process many times the intensity and phase-space density can be raised to the instability limit for the machine. The usual focusing and accelerating fields in accelerators (i.e. those derived from a Hamiltonian) cannot merge two sets of identical particles into the same phase space. This fact is enshrined in the Liouville Theorem [22]. Thus, the charge-exchange injection is often referred to as a way of cheating Liouville. In order that the circulating beam does not permanently traverse the stripping foil, either the foil must be removable or injection is carried out with a closed-orbit bump (see Fig. 8). One side of the foil must be unsupported, or the support would destroy the beam as the injection bump is removed or the foil is withdrawn. The stripping foil can be made from aluminium oxide [23].

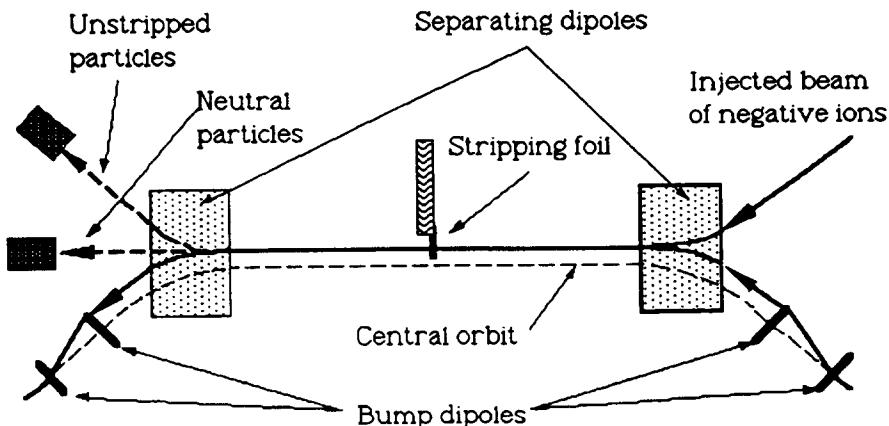


Fig. 8 H<sup>-</sup> injection insertion

## 6. COLLIMATION INSERTIONS

High-efficiency collimation is becoming more and more important. The aims are to minimise background in physics detectors, avoid quenching magnets in superconducting machines and to limit the irradiation of equipment in transfer lines. For these exacting tasks, the simple concept of placing two metal blocks either side of the beam to define an aperture is no longer sufficient. Three high-efficiency insertions will be analysed for rings and one for a transfer line;

- 1-D system with zero dispersion (betatron collimation)
- 1-D system with dispersion but zero bending (betatron & momentum collimation)
- 2-D system with dispersion, but zero bending (betatron & momentum collimation)
- 2-D system with zero dispersion for transfer lines (betatron collimation).

It is convenient to use normalised variables when analysing the action of the collimators so the main equations are developed below. The familiar parametric form of a betatron oscillation is,

$$y(s) = F\sqrt{\beta(s)} \cos(\mu(s) + \phi) \quad (47)$$

where  $y(s)$  is either of the transverse coordinates  $x$  or  $z$  and  $F$  and  $\phi$  are constants. The normalised phase-space coordinates are then defined as,

$$Y(\mu) = \left( \frac{y}{\sqrt{\beta}} \right) = F \cos(\mu + \phi) \quad (48)$$

and

$$\frac{dY}{d\mu} = \frac{d}{d\mu} \left( \frac{y}{\sqrt{\beta}} \right) = -F \sin(\mu + \phi) \quad (49)$$

where  $Y$  would be replaced by  $X$  or  $Z$  when appropriate. This transformation into normalised phase space reduces the betatron oscillations to simple harmonic motion. For later use, it is necessary to express  $(dY/d\mu)$  in terms of more usual functions, that is,

$$\frac{dY}{d\mu} = \frac{d}{d\mu} \left( \frac{y}{\sqrt{\beta}} \right) = \frac{1}{\sqrt{\beta}} \frac{dy}{d\mu} - \frac{1}{2} \frac{y}{\beta^{3/2}} \frac{d\beta}{d\mu} \quad (50)$$

but

$$d\mu = \frac{1}{\beta} ds \quad \text{and} \quad \alpha = -\frac{1}{2} \frac{d\beta}{ds} \quad (51)$$

so that

$$\frac{dY}{d\mu} = \sqrt{\beta} \frac{dy}{ds} + \alpha \frac{y}{\sqrt{\beta}} . \quad (52)$$

In a bending-free region, the dispersion function  $D(s)$  behaves exactly like a betatron oscillation. In analogy with the betatron case above, the dispersion oscillation and the corresponding normalised dispersion variables ( $\chi$ ,  $d\chi/d\mu$ ) can then be defined as,

$$D(s) = D_0 \sqrt{\beta(s)} \cos(\mu(s) + \phi) \quad (53)$$

$$\chi = \frac{D}{\sqrt{\beta}} = D_0 \cos(\mu + \phi) \quad (54)$$

and

$$\frac{d\chi}{d\mu} = \sqrt{\beta} \frac{dD}{ds} + \alpha \frac{D}{\sqrt{\beta}} = -D_0 \sin(\mu + \phi) . \quad (55)$$

It is important to note the similarity between betatron and dispersion oscillations and in particular, their complete equivalence in a bending-free region, since this is the underlying property that will make it possible to combine betatron and momentum collimation. This can be understood in general terms from the differential equations for particle motion in an alternating gradient lattice [3]. The motion due to a deviation in momentum from the nominal value is derived by adding a forcing term of the form  $(1/\rho)(\Delta p/p)$  to the basic equation, but in a bending-free region  $\rho = \infty$ , the forcing term becomes zero and the equation reverts to its non-dispersive form. Thus in a bending-free region, a dispersion oscillation has exactly the same form and properties as a betatron oscillation (to first order).

Figure 9 shows a normalised phase-space plot of a beam with a dense low-emittance core and a diffuse high-emittance momentum tail. Some further explanation is needed to explain the combination of the betatron and dispersion oscillations in phase space. The betatron oscillations are not closed over one turn and fill out an elliptical phase space area. The

dispersion oscillation, on the other hand, is a closed oscillation and, at a given azimuth in the machine, it defines only a point. The position coordinate of this point varies with the momentum deviation and is the reason for the radial distribution of the ellipses in Fig. 9. In a similar way, the ellipses are also spread out parallel to the vertical axis to take into account the longitudinal angle coordinate of the dispersion trajectory. The latter can be omitted in most cases, since a movement parallel to the image of a collimator in phase space has no effect on the interception of particles, but it will be necessary to take this angle into account, when the ideal phase of the dispersion oscillation is specified for a combined collimation scheme and also when aligning the collimator.

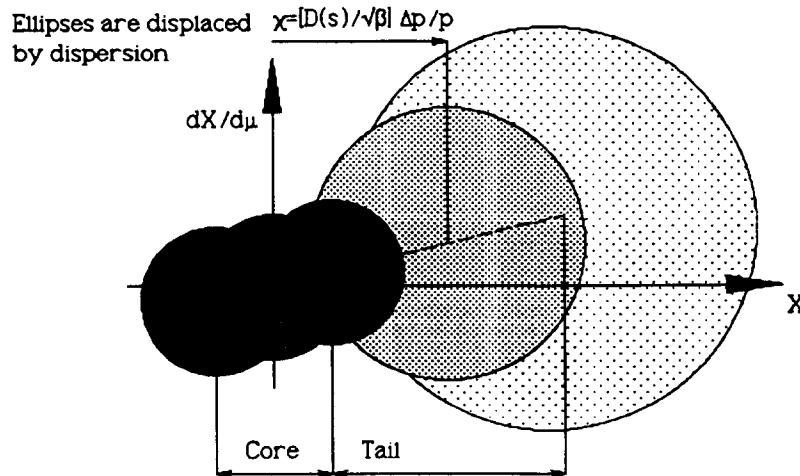


Fig. 9 Normalised phase-space plots for a beam with a momentum tail

### 6.1 One-dimensional betatron collimation

For pure betatron collimation, the dispersion is zero, so that all emittance ellipses coincide and all momenta are treated equally (see Fig. 10).

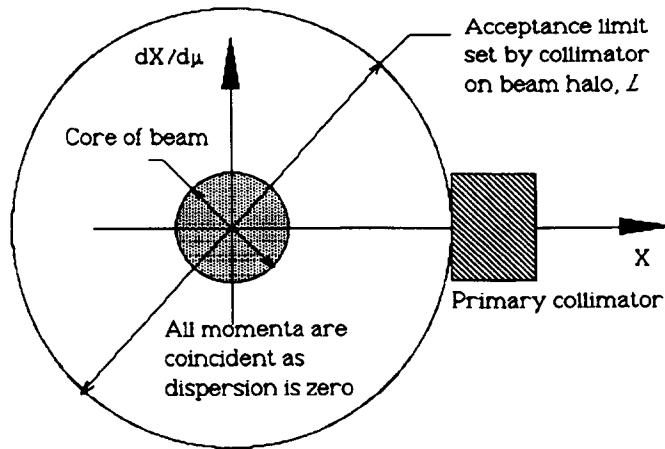


Fig. 10 Action of primary collimator when dispersion is zero

The primary collimator is followed by two secondary collimators set at optimised phases for intercepting the scattered particles. The principle for positioning the secondary collimators is thought to have been first given by Teng in [24]. It is independently developed in [25, 26] and it has been applied in [27]. Since the scheme is sited in a non-dispersive region any energy loss suffered by these scattered particles in the primary collimator is unimportant (to first order). The main features of the scheme are shown in Fig. 11.

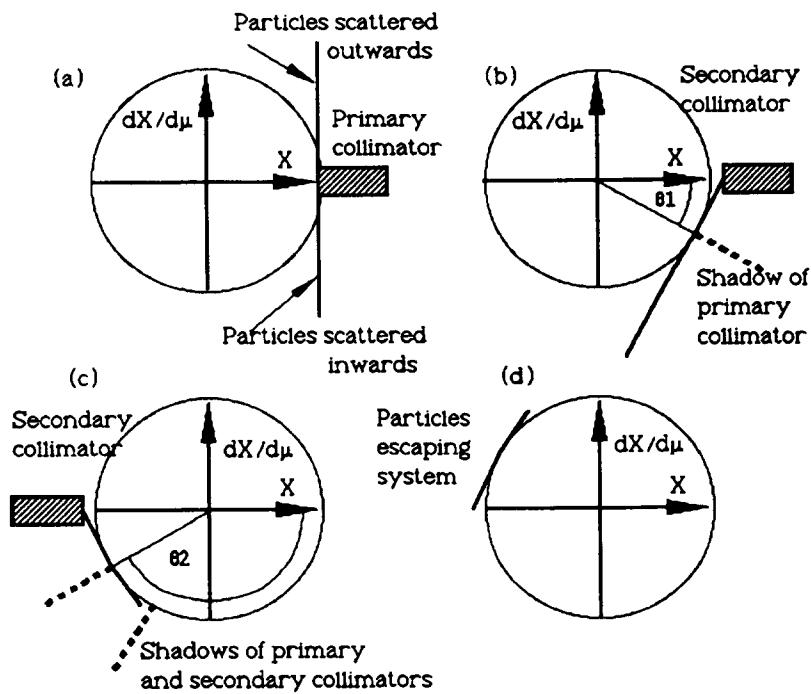


Fig. 11 Main features of a single-stage betatron collimation system

The optimum phase shifts between the primary collimator and its secondaries (see  $\theta_1$  &  $\theta_2$  in Fig. 11) will correspond to the most efficient interception of the scattered 'tails'. Consider Fig. 12, in which the situation in Fig. 11 (b) is shown in more detail. The scattered particles are spread out along a line at an angle  $\theta$  equal to the betatron phase advance between the primary and secondary collimators. The secondary collimator is stepped back from the beam envelope to prevent it becoming a new primary collimator. There is an optimum  $\theta$  for which the length of the line PR of the locus of the scattered particles is a minimum. This optimum depends upon the radius of the beam envelope, which equals the half-aperture of the primary collimator  $L$  (normalised) and the step back  $H$  (normalised).

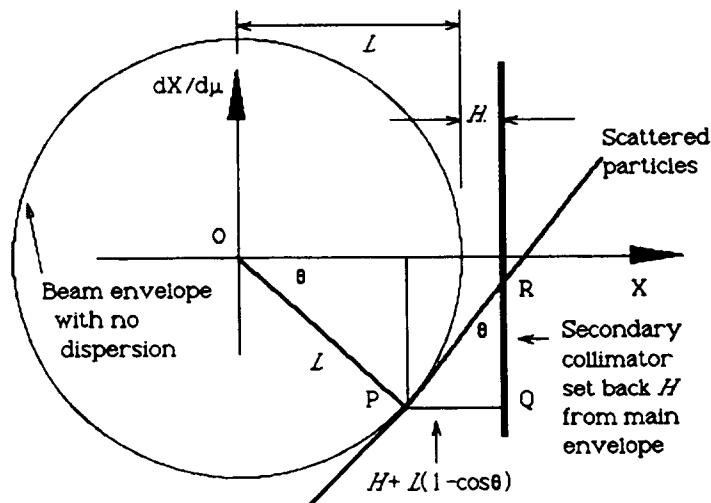


Fig. 12 Action of secondary collimator [Detail of Fig. 11(b)]

By geometry,

$$PR = \frac{[H + L(1 - \cos \theta)]}{\sin \theta} \quad (56A)$$

and after rearranging

$$PR = L \tan\left(\frac{\theta}{2}\right) + \frac{H}{\sin\theta} . \quad (56B)$$

As a quick check it can be seen that  $PR \rightarrow \infty$  when  $\theta = 0$  and  $PR = L+H$  when  $\theta = 90^0$ . The minimum of PR is found by differentiation of (56B),

$$\frac{d PR}{d\theta} = \frac{L}{2} \sec^2\left(\frac{\theta}{2}\right) - \frac{H \cos\theta}{\sin^2\theta} = 0 \quad (57)$$

which yields the optimum phase advance  $\theta_1$  of the first secondary collimator as,

$$\theta_1 = 2 \sin^{-1} \sqrt{\left[ \frac{1}{2} \left( \frac{H}{L+H} \right) \right]} . \quad (58)$$

By inspection, it can be seen that the phase advance  $\theta_2$  of the second secondary collimator will be  $180^0 - \theta_1$ . Equation (58) has been evaluated numerically in Table 3.

It can be seen from Fig. 12 that the particles escaping the system are contained within a normalised acceptance circle of radius,  $OR^2 = PR^2 + L^2$  and a little further manipulation will show that at the optimum phase advance  $\theta_1$ , the point R moves onto the real axis, so that OR is a minimum and is simply given by,

$$OR_{\min} = L + H . \quad (59)$$

If this is still within the acceptance of the rest of the machine, then the escaping particles will be trapped on the primary collimator some turns later and the overall efficiency is improved. Table 3 includes the normalised acceptance of the escaping particles.

**Table 3**  
Typical parameters for a single-stage collimation scheme

H/L	0	0.05	0.1	0.15	0.2	0.25
$\theta_1$ [degree]	0	17.8	24.6	29.6	33.6	36.9
OR <sub>min</sub> *	L	1.05L	1.1L	1.15L	1.2L	1.25L

\* The limit acceptance OR of the escaping particles is expressed in units of L which is the normalised half-aperture of the primary collimator. A practical unit for L and OR that is used widely for apertures would be a standard deviation of the nominal beam emittance.

There is a complication in the above model. The primary collimator also scatters the escaping particles in the orthogonal plane and, in many cases, the increase in amplitude will be far larger than in the plane of collimation causing the particles to be lost elsewhere in the machine. This problem will be confronted in the third collimation insertion.

## 6.2 One-dimensional betatron and momentum collimation [28]

In the presence of dispersion, the primary collimator will no longer cut the betatron emittance ellipses for all momenta to the same degree. The primary collimator will impose a different acceptance ellipse for each momentum as shown in Fig. 13. This figure is the sequel to Fig. 10 and shows the beam at the exit of the collimator (for the most general case the ellipses in Fig. 13 would also be spread out vertically, but to limit the number of figures the final result is preempted and the peak of the dispersion oscillation has been chosen).

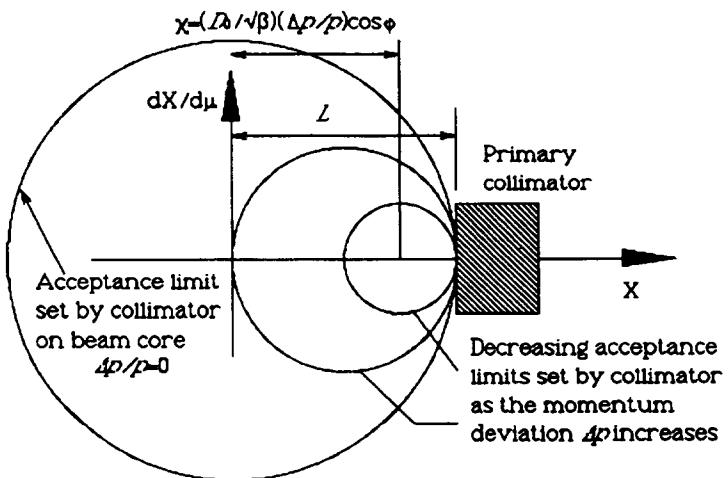


Fig. 13 Action of primary collimator when dispersion is non-zero

If  $A$  is defined as the normalised amplitude of the betatron acceptance ellipse for a given momentum deviation, then the primary collimator will set this limiting amplitude according to,

$$A = L - \frac{D_0}{\sqrt{\beta}} \frac{\Delta p}{p} \cos \varphi \quad (60)$$

where  $L$  is the normalised half-aperture of the primary collimator and  $\mu$  is taken as zero at the primary collimator so that  $\varphi$  gives the phase of the dispersion oscillation.

At a betatron phase  $\theta$  later, the beam will pass the first of the two secondary collimators that are set back from the beam envelope by a distance  $H$ . It is assumed that the beam envelope is defined by the acceptance ellipse set by the primary collimator on the core beam of nominal momentum and this assumption will be justified later. Upon arrival at the secondary collimator, all ellipses will have rotated by  $\theta$ , but those with a non-zero momentum deviation will also have been moved radially by the dispersion oscillation. This is illustrated in Fig. 14 for an individual off-momentum ellipse that is moved towards the central orbit and away from the collimator by the dispersion oscillation.

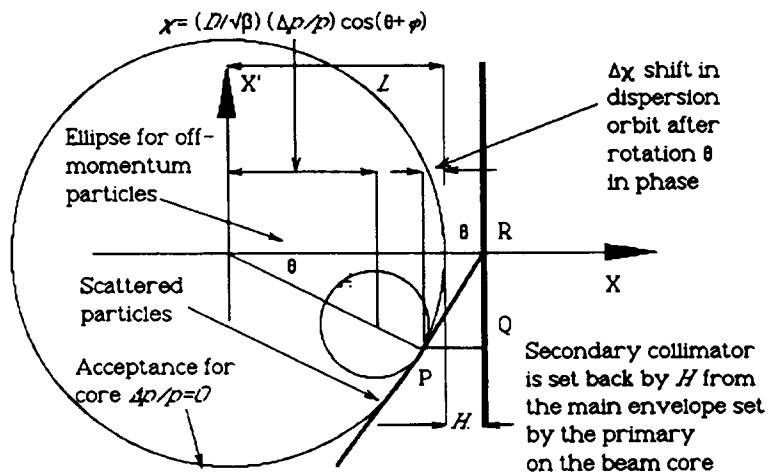


Fig. 14 Action of a secondary collimator with dispersion

For the off-momentum ellipse the scattered particles lie along the line PR at the time the secondary collimator is passed. The length of the line PR represents the particles that escape

the collimator and is therefore a measure of the system's efficiency. Clearly, an uncontrolled dispersion is potentially very dangerous since it could move the off-momentum ellipse away from the central orbit. From the geometry of Fig. 14,

$$PR = PQ / \sin \theta = [H + \Delta\chi + A(1 - \cos \theta)] / \sin \theta \quad (61)$$

where  $\Delta\chi$  is the normalised movement of the dispersion orbit and is given by,

$$\Delta\chi = \frac{D_0}{\sqrt{\beta}} \frac{\Delta p}{p} [\cos \varphi - \cos(\varphi + \theta)] . \quad (62)$$

If the phase of the normalised dispersion oscillation is chosen to be zero at the primary collimator (i.e.  $\varphi = 0$ ), then

$$\Delta\chi = \frac{D_0}{\sqrt{\beta}} \frac{\Delta p}{p} [1 - \cos \theta] . \quad (63)$$

The length of PR given in (61) can now be written as,

$$PR = [H + \frac{D_0}{\sqrt{\beta}} \frac{\Delta p}{p} (1 - \cos \theta) + A(1 - \cos \theta)] / \sin \theta \quad (64)$$

but  $A$  has already been determined in (60), so by substitution into (64),

$$PR = \frac{[H + L(1 - \cos \theta)]}{\sin \theta} . \quad (65)$$

This is exactly the same result as was obtained in (56A) for the non-dispersion case and consequently the optimum phase advances for the secondary collimators remain unchanged. In theory, chromatic effects will adversely affect the above by making  $\theta$  a function of momentum, but in practice, a collimation insertion will have less than  $180^\circ$  phase advance and it is unlikely that sextupoles will be included, so these complications can be ignored. Three important statements can now be made :

- (a) When the normalised dispersion has its peak at the primary collimator, all momenta are treated equally and require the same standard phases for the secondary collimators.

The condition for the normalised dispersion oscillation to have its peak at the primary collimator can be found by equating (55) to zero,

$$\frac{d\chi}{d\mu} = \sqrt{\beta} \frac{dD}{ds} + \alpha \frac{D}{\sqrt{\beta}} = 0 \quad (66)$$

to give the collimation condition ,

$$\frac{1}{D} \frac{dD}{ds} = -\frac{\alpha}{\beta} \text{ at the primary.} \quad (67)$$

- (b) The longitudinal slope of the beam is the same for all momenta, so the alignment of the primary collimator will have the same efficiency as in the betatron case.

The amplitude of the beam envelope at the entry to the collimator is given by the contributions of two terms, one coming from the betatron oscillation envelope in (47), the other coming from the dispersion oscillation, which is maximum at this point.

$$x_{\text{collimator}} = F\sqrt{\beta} + D \Delta p / p \quad (68)$$

After differentiation,

$$\begin{aligned} x'_{\text{collimator}} &= \frac{1}{2} \frac{F\sqrt{\beta}}{\beta} \frac{d\beta}{ds} + \frac{dD}{ds} \frac{\Delta p}{p} \\ &= -\frac{\alpha F\sqrt{\beta}}{\beta} + \frac{dD}{ds} \frac{\Delta p}{p}. \end{aligned} \quad (69)$$

Taking into account the collimation relation (67),  $\frac{dD}{ds} = -\frac{\alpha}{\beta} D$

one obtains with (68)

$$x'_{\text{collimator}} = -\frac{\alpha F\sqrt{\beta}}{\beta} + \left( -\frac{\alpha}{\beta} D \right) \frac{\Delta p}{p} = -\frac{\alpha}{\beta} x_{\text{collimator}} \quad (70)$$

which is independent of momentum.

- (c) As far as the efficiency (i.e. capture of particles) is concerned, there is no preferred value for the dispersion amplitude  $D_\theta$ .

Thus a combined betatron and momentum collimation scheme requires the condition (67) to be fulfilled at the primary collimator, to ensure that the dispersion and betatron oscillations have their correct phases. In the past, when schemes have been installed in dispersive regions, they have either been positioned by other criteria as in the ISR [29], or by simulations as in LEP [30].

### 6.3 Two-dimesional betatron and momentum collimation [28]

The main drawback of the systems described so far is the lack of control of the scattering in the orthogonal plane. This can be solved by working in two-dimensions with a shaped primary collimator.

#### 6.3.1 More about the beam model

In phase space the beam particles follow ellipses or circles, whereas in real space they sweep over rectangular areas as shown in Fig. 15(a). The motion is close to being simple harmonic in the two planes so that the probability density of finding a particle in its rectangle is highest in the corners and lowest at the centre. This assumes that the particles are not locked in resonances, but have non-rational tunes. The largest rectangles that can exist within the aperture will just touch the aperture limit. In this way, a beam takes up the shape of an aperture limit and fills the space. The fact that beams always appear dense at the centre and elliptical, or circular, in shape may appear to refute this, but this effect is due to the very large number of particles within a normal beam and the way their 'rectangles' are distributed by size and aspect ratio.

This model will be used to compare the scattering actions of a collimator that is parallel to one of the betatron oscillation axes to one that is inclined. It will also give an intuitive indication of the results obtained in [31] for the impact parameter.

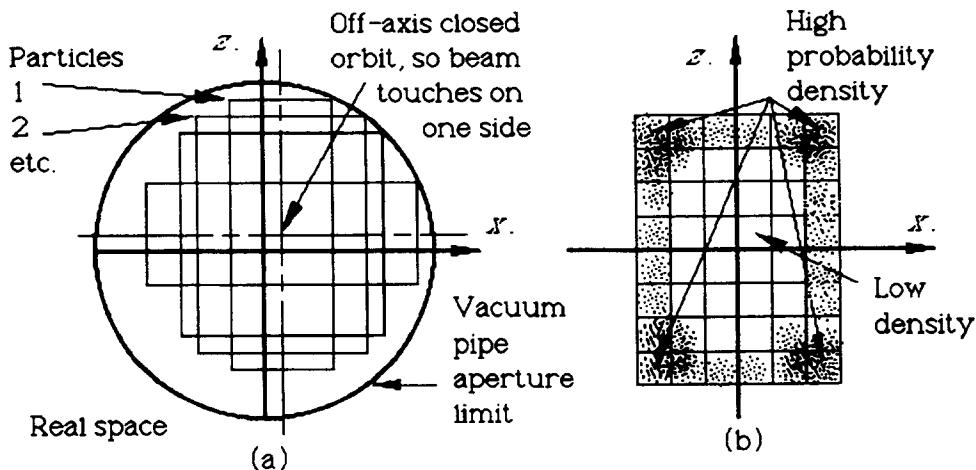


Fig. 15 Beam model in real space.  
 (a) Some particles with an off-axis closed orbit at the aperture limit.  
 (b) Probability density swept out by one particle (schematic).

### 6.3.2 Scattering in the orthogonal plane [31]

First consider scattering in an upright collimator aligned with the vertical betatron oscillations as shown in Fig. 16. When a particle hits the collimator it is scattered without any preferential direction, so it is necessary to consider scattering in the plane of collimation and in the orthogonal plane.

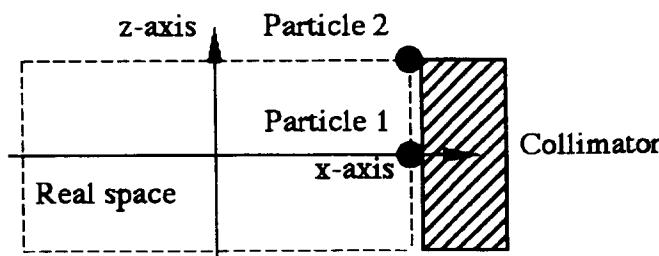


Fig. 16 Upright collimator against a beam in real space

In Fig. 16, particle 1 on the  $x$ -axis is at the peak of its horizontal oscillation and at the centre of its vertical oscillation, while particle 2 is at the peak of both its horizontal and vertical oscillations. For argument's sake, it is assumed that the scattering angles are equal in both planes and for both particles. The normalised phase-space plots then appear as shown in Fig. 17.

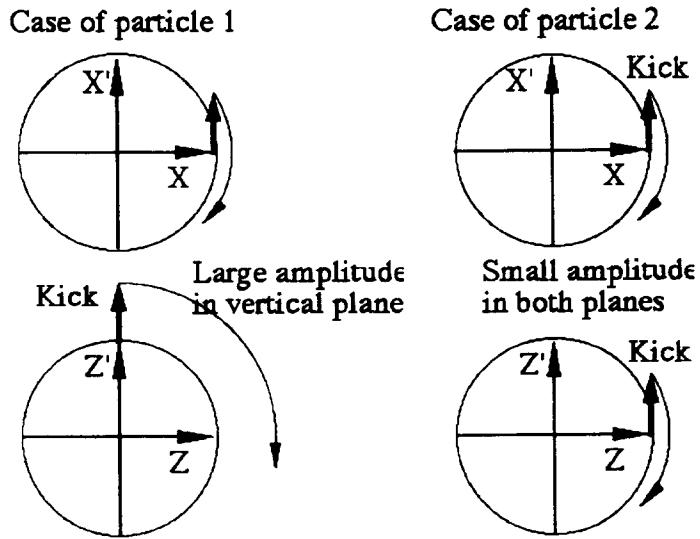


Fig. 17 Normalised phase-space plots for scattered particles

Horizontally the situation is equal for the two particles. The scattering causes a shift parallel to the  $X'$ -axis and the particle moves in phase but little in amplitude. This is the case described in Section 6.1. In the vertical plane, particle 2 is also at the maximum of its vertical oscillation and its phase-space diagram is a repeat of the horizontal plane. Particle 1 is different, however, and the kick adds fully to its amplitude. In Section 6.1, it is explained that particles with very small horizontal scattering will escape the secondary collimators. These escaping particles will have a distribution of vertical amplitudes similar to particle 1 and are likely to be lost later in the machine. This is the problem of uncontrolled scattering in the orthogonal plane.

There is an advantage in minimising the particles of type 1, since unlike particles of type 2, their enhanced amplitude gives them little chance of surviving a full turn of the machine and being caught by the collimator system some time later. A collimator with an inclined, or curved face, such as the circular aperture limit shown in Fig. 15(a) will preferentially select the particles that are at the peaks of both their oscillations and will ensure controlled scattering in both planes.

### 6.3.3 Insertion optics and design of a 2-dimensional collimator system

The design of a 2-dimensional system is made easy by choosing equal phase advances with distance in both planes and equal betatron amplitudes. This can be achieved by adopting a low- $\beta$  optics as shown in Fig. 18. This has already been proposed for LHC in Ref. [26] for pure betatron collimation with two 1-dimensional systems.

Into this optics place a primary collimator with a circular aperture  $L$  followed by a secondary collimator with a circular aperture  $L+H$ . The shape of the primary collimator aperture will preferentially select particles at the maxima of their oscillations and a typical particle with a momentum deviation is shown in Fig. 19. In the secondary collimator, the position of the unscattered particle is shown with a schematic circle indicating the possible scattering.

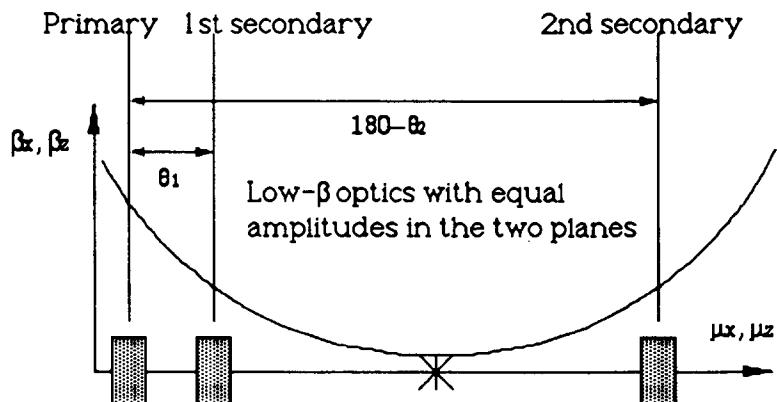


Fig. 18 Optics for a 2-dimensional collimator system

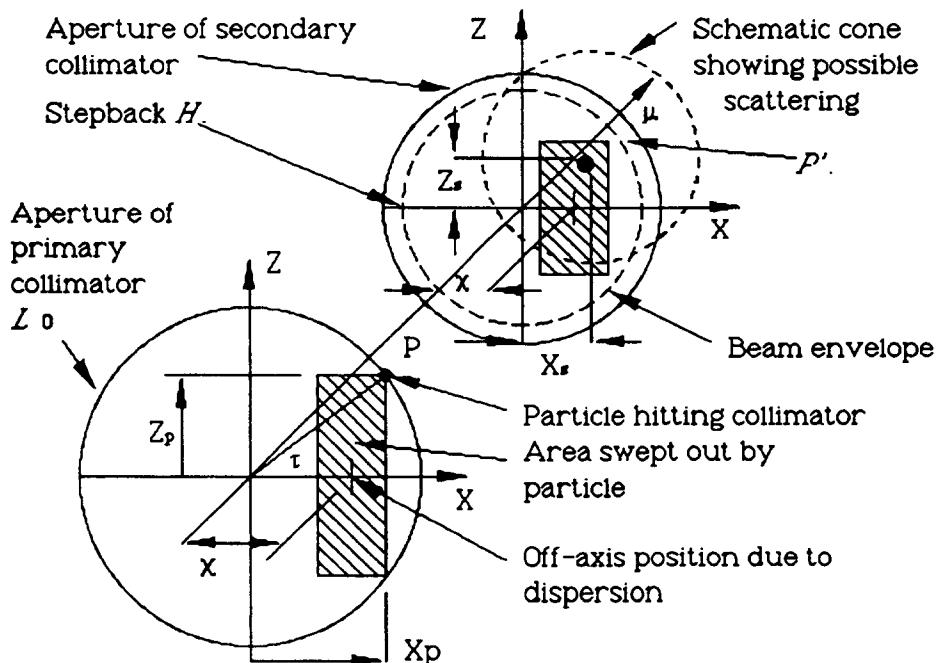


Fig. 19 Action of primary and secondary collimators in a two-dimensional scheme

The beam envelope at  $P$  maps to  $P'$ , which in real space scales with  $\sqrt{\beta_x}$  and  $\sqrt{\beta_z}$  but is invariant in normalised coordinates. The secondary collimator is placed outside this point as shown and its exact shape will be derived below. The particle at  $P$  with angle  $\tau$  has been limited to the horizontal and vertical apertures,

$$X_p = L_0 \cos \tau \quad \text{and} \quad Z_p = L_0 \sin \tau \quad (71)$$

where  $L_0$  is the normalised radius of the primary aperture and  $\tau$  can assume any value from  $0 \rightarrow \pi/2$ . This is equivalent to placing normal collimators at  $X_p$  and  $Z_p$  in the usual way. The  $X$  and  $Z$  motions between the collimators can be treated independently and correspond to all the previous results with  $L$  replaced by (71). Since the betatron amplitudes have been chosen as equal in the two planes both the horizontal and vertical motions require the secondary collimator at the same place. Figure 14 then applies in both planes, except that in the vertical plane the dispersion is zero and it degenerates to Fig. 12. Now place hypothetical collimators of the normal type at the coordinates of the point  $R$  in real space. It was shown in Section 6.1 in (59)

that the point R would be at the normalised coordinates  $(L+H)$  in each plane. If now the point P is moved to anywhere on the aperture limit of the primary collimator the locus of R will describe the profile of the secondary collimator. With the use of (71),

$$\begin{aligned} OR_{\min}^2 &= (L+H)_{\text{HORIZ}}^2 + (L+H)_{\text{VERT}}^2 \\ &= (L_0 \cos \tau + H)^2 + (L_0 \sin \tau + H)^2 \\ &= L_0^2 + 2H^2 + 2\sqrt{2}L_0H\sin(\tau + \pi/4). \end{aligned} \quad (72)$$

The profile given by (72) for  $0 \leq \tau \leq \pi/2$  should be applied to each quadrant. When  $\tau = 0$  or  $\pi/2$ ,  $OR_{\min} = L_0 + H$  (to first order in  $H/L$ ), which is the normal 1-dimensional situation. When  $\tau = \pi/4$ ,  $OR_{\min} = L + \sqrt{2}H$ . Thus, there is a small correction to be made to the surface of the secondary collimators that increases the standard stepback  $H$  on the X- and Z- axes to  $\sqrt{2}H$  on the  $45^\circ$  lines.

Furthermore, if  $L_0$  is made a function of  $\tau$  then (72) can be used to find the profile of the secondary for any shape of primary.

$$OR_{\min} \approx [L_0(\tau) + H] \left\{ 1 - [H / L_0(\tau)] [1 - \sqrt{2} \sin(\tau + \pi/4)] \right\}. \quad (73)$$

All the results of the previous sections can now be applied directly in the 2-dimensional case. The optimum secondary collimator phases are equal in the two planes and are given by (58). If dispersion is required in one plane then the condition (67) must be applied at the primary collimator. Finally, the profile of the secondary is given by (73) where  $L_0(\tau)$  is the profile of the primary.

The above scheme has the major advantage over the preceding schemes that any particle that escapes the system will be contained within the normalised aperture  $OR_{\min} = L+H$  in both planes (see Section 6.1 and Table 3). If therefore, the primary collimator is set so that the machine acceptance is greater than, or equal to,  $L+H$  in both planes, then the efficiency of the insertion can theoretically reach 100 percent assuming that all particles that hit a secondary collimator are absorbed. More collimators could be added to intercept even these particles, but there would be a penalty to pay in the time needed to set up and optimise the system.

### 6.3.4 Two-dimensional collimator design

In an unprotected machine, the circular vacuum chamber would act somewhat like the 2-dimensional collimator described above. Thus, for chamber protection, circular primary and secondary collimators would appear to be a good choice, which would also be insensitive to coupling as mentioned in [28]. Inclined collimators set at  $45^\circ$  would also be possible. In practice, a hybrid of these two possibilities is the simplest to build as shown in Fig. 20.

### 6.3.5 Enhancement of the impact parameter

Figure 15(b) gives a schematic indication of the probability density to find a particle in real space. An upright collimator aligned with one of the betatron oscillations will integrate over the probabilities along one full side of the rectangle, whereas an inclined collimator can only intercept the corner of the rectangle. It is expected therefore that more turns are needed on average to hit an inclined collimator than an upright one. If the beam is slowly expanding, or the collimator advancing, then the extra turns should enhance the impact parameter. This is explored in more detail in [28].

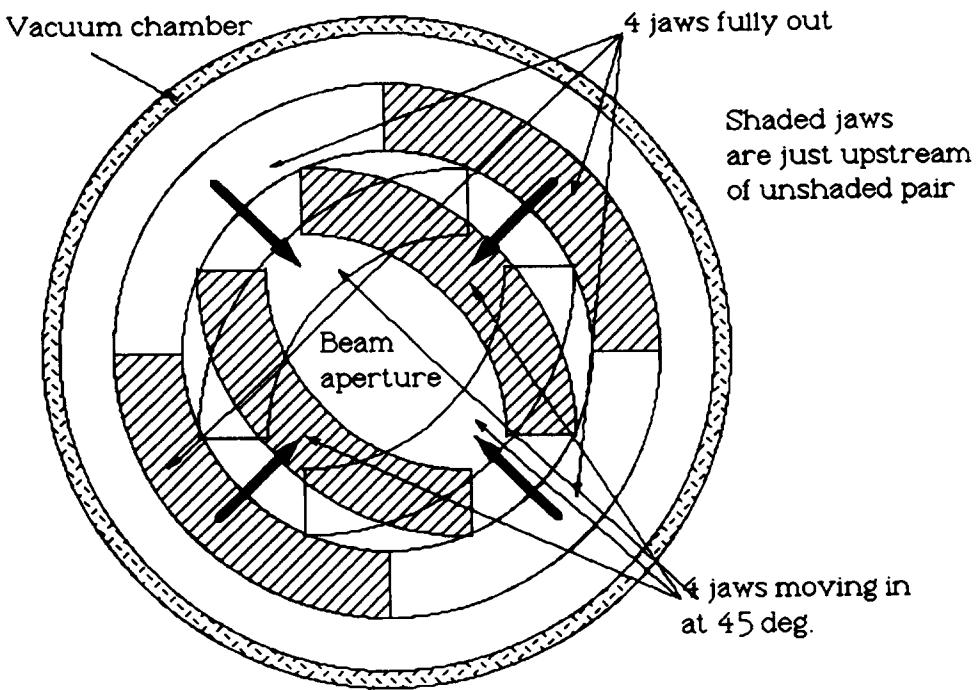


Fig. 20 4-jaw circular collimator

#### 6.4 Collimation systems in regions with bending

Although collimation in the presence of dispersion has been analysed, it was assumed that the collimation insertion itself was situated in a bending-free region, so that the dispersion vector behaved as a betatron oscillation. In many cases, this may not be possible to arrange and it is useful to know how to adapt the insertions described here to this more general situation. The mathematical tools necessary for doing this are described in Ref. [32].

#### 6.5 Two-dimensional betatron collimation insertion for a transfer line

Although the same basic principles apply, the action of the primary collimator(s) is a little different in a transfer line compared to a ring. In a ring, the periodic return of the beam means that only one primary collimator is needed to select the particles at the maxima of their amplitudes, whereas in a transfer line a series of collimators is needed. This makes it necessary to extend the system and to make some small compromises.

One possible solution would be to use a dispersion-free FODO lattice with a phase advance per cell of  $72^\circ$ . Position the first 2-dimensional primary (with circular jaws on all sides of the beam) at the point midway between F and D units where the betatron amplitudes are equal in the two planes. The secondaries for this primary can be placed two and three cells later at  $144^\circ$  and  $216^\circ$ . These points will also be midway between F and D units, the betatron amplitudes will be equal and the phase advances the same in the two planes [see Fig. 21(a)]. These positions are at  $180^\circ \pm \theta_1$ , which determines  $\theta_1$  as  $36^\circ$  and the optimum stepback as  $H/L = 0.25$ . Note that these positions are fully equivalent to the  $0^\circ + \theta_1$  and  $180^\circ - \theta_1$  used in the earlier examples (see Fig. 11). It is necessary to repeat this system at three consecutive points midway between F and D units as shown in Figs. 21(a) to (c). This is necessary in order to trap those particles whose phases enabled them to escape the first collimator. Where a primary falls on the position of a secondary, the primary takes precedence and acts equally as the secondary. The resultant collimation array is shown in Fig. 21(d). The primary collimators cut the phase space of the beam as shown in Fig. 21(e). The irregular polygon made by the

collimators has an inscribed circle of radius  $L$ , but in the corners particles can survive with larger amplitudes up to  $L/\cos(\theta_1/2)$ , which is  $1.24L$ . This matches almost exactly the amplitude  $(L+H) = 1.25L$  that will survive the two-dimensional collimator array. Thus, this system will collimate the beam to a normalised amplitude that is just 25% bigger than the primary aperture.

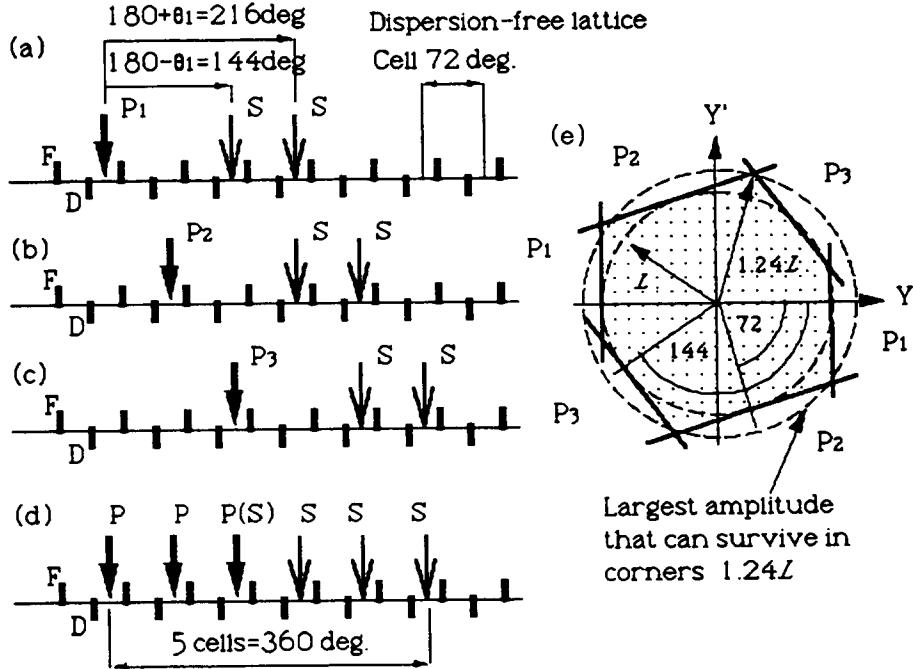


Fig. 21 A betatron collimation insertion for a transfer line  
 (a) Basic scheme, (b) and (c) Repeat of basic scheme at  $72^\circ$  and  $144^\circ$ ,  
 (d) Full collimator array, (e) Action of primaries in phase space.

## 7. MORE EXOTIC INSERTIONS

### 7.1 Zero-angle crossing in two-ring colliders

The introduction of a zero-angle crossing scheme into a two-ring collider is desirable for high luminosity and minimisation of the beam-beam effect. On the other hand, it adds to the complication of the lattice design and introduces common elements that limit the operational flexibility of the machine. For example, running with asymmetric energies is made difficult, if not impossible, when there are common dipoles. Three crossing schemes will be considered.

When the same particle is used in both rings, common dipoles can be used to bring the orbits together into head-on collision and then to separate them again [see Fig. 22(a)]. The low- $\beta$  quadrupoles can then be mounted in the common section, but this does have the disadvantage that small bunch spacings will lead to unwanted crossings within the insertion. In LHC, this has led to the compromise of the above geometry with a micro-crossing angle of  $200\mu\text{rad}$  [26].

The electron-positron machines cannot separate their beams using common dipoles, but the same layout can be achieved using common quadrupoles with a small crossing angle [see Fig. 22(c)], or with electrostatic separators [see Fig. 22(b)]. In modern designs, the two rings may be stacked vertically, so that the crossing scheme works in the vertical plane, while the normal dispersion suppression and guide field act horizontally. Controlling both the horizontal

and vertical dispersion is a further complication in the optics. This can be solved by placing the vertical bends inside the horizontal-dispersion-free region and arranging the vertical dispersion as a full-wave bump with zero amplitude at the crossing point.

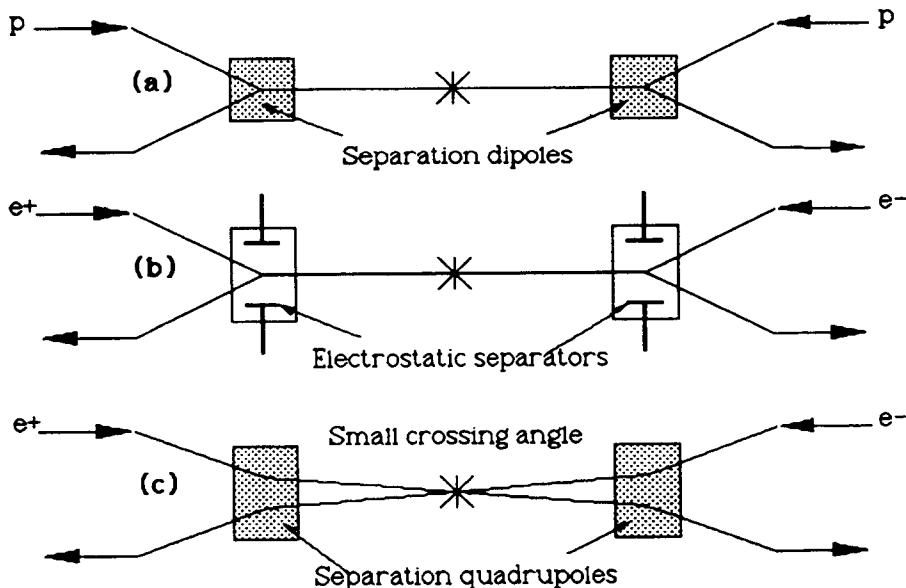


Fig. 22 Zero- and small-angle crossings

## 7.2 Crab crossing

The 'crab crossing' is an astute, but as yet untried, scheme [33] for enjoying the advantages of head-on collision while maintaining a crossing angle (see Fig. 23). It avoids common elements and the deflection of the beam close to the interaction point, which reduces the amount of background radiation entering the experiment. The scheme was first proposed for linear colliders, but is actively studied for the high-intensity, electron-positron colliders.

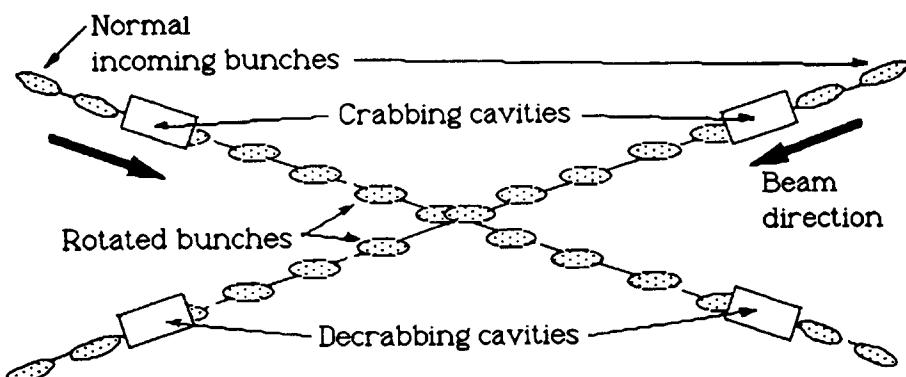


Fig. 23 Crab crossing

Just before the bunches reach the crossing point, they are tilted by half the crossing angle with a strong rf cavity, so that they cross the interaction region in a crab-like posture. The bunch crossing then takes place head-on, but with a coherent sideways movement. The bunches are then rotated back to the axis before making the next turn of the machine.

### 7.3 Monochromator

Bunched beams inevitably have a momentum spread and in the bunch-bunch collision there is some uncertainty in the energy of the collision and hence the energy available for particle production. The 'monochromator' attempts to reduce this uncertainty by arranging an antisymmetry between the two rings for the variation of transverse position with momentum at the crossing point. In this way, the high momentum particles of one beam are made to collide with the low momentum particles of the other and vice versa. This idea has been discussed for some time, but it still awaits a practical demonstration of its advantages.

### 7.4 Insertion devices

In many cases the beams in a storage ring or collider are passed through equipment that would not normally be part of a synchrotron, for example, undulators in synchrotron light sources, solenoids in colliders, and electron coolers in small storage rings. In these examples, there is some influence on the synchrotron and special optics and correction elements have to be provided in an insertion around the device. For electron cooling, the optics design employs the basic layout of the  $H^-$  injection in Fig. 8, except that there is no orbit bump, no stripping foil and the optics of the electron beam is more critical and an axial magnetic field has to be applied to keep it confined. An excellent discussion of the theory and design of electron cooling insertions can be found in Ref. [34].

\* \* \*

## REFERENCES

- [1] E.D. Courant, M.S. Livingston, H.S. Snyder, "The strong-focusing synchrotron - a new high energy accelerator", Phys. Rev., 88, (1952), 1190-6., and E.D. Courant, H.S. Snyder, "Theory of the alternating gradient synchrotron", Annals of Physics, 3, (1958), 1-48.
- [2] N.C. Christofilos, unpublished report (1950), and US Patent No. 2.736,799, filed March 10, 1950, issued February 28, 1956.
- [3] K. Steffen, "Basic course on accelerator optics", Proc. CERN Accelerator School General Accelerator Physics Course, Gif-sur-Yvette, 1984, CERN 85-19, (1985), 25-63, or alternatively,  
P. Schümser, "Basic course in accelerator optics", Proc. CERN Accelerator School Second General Accelerator Physics Course, Aarhus, 1986, CERN 87-10, (1987), 1-44.
- [4] E. Regenstreif, "Phase-space transformations by means of quadrupole multiplets", CERN 67-6, (1967).
- [5] E. Regenstreif, "Possible and impossible phase-space transformations by means of alternating-gradient doublets and triplets", CERN 67-8 (1967).
- [6] G. Wüstefeld, "Analytic beta function matching with the thin-lens doublet", Proc. CERN Accelerator School Third General Accelerator Physics Course, Salamanca, 1988, CERN 89-05, (1989), 102-4.
- [7] B. Zotter, "Calculation of the parameters of a quadrupole triplet from the transfer matrices", CERN/ISR-TH/73-43, (1973).
- [8] M.D. Woodley, M.J. Lee, J. Jäger, A.S. King, "COMFORT", IEEE Trans. Nucl. Sci., NS-30, (1983), 2367.
- [9] R. Servranckx, K.L. Brown, "DIMAD", SLAC Report 270 UC-28, (1984).
- [10] F.C. Iselin, H.G. Grote, "The MAD Program", Version 8.4, CERN/SLL/90-13, (1991).
- [11] K. Steffen, J. Kewisch, "PETROS", DESY PET 76/09, (1976).
- [12] G. Guignard, Y. Marti, "PETROC Users' Guide", CERN ISR-BOM-TH/81-32, (1981).
- [13] A.A. Garren, A.S. Kenney, E.D. Courant, M.J. Syphers, "SYNCH", Fermilab Report FN 420, (1985).
- [14] L. Schachinger, R. Talman, "TEAPOT", SSC-52, (1985).

- [15] R. Brinkmann, "Insertions", Proc. CERN Accelerator School Second General Accelerator Physics Course, Aarhus, 1986, CERN 87-10, (1987), 45–61.
- [16] H.B. Dwight, "Tables of integrals and other mathematical data", 4th edition, (The Macmillan Co., New York, 1961), 92.
- [17] K. Robinson, G.A. Voss, "Operation of a synchrotron as a storage ring", Symposium international sur les anneaux de collisions, Saclay, 1966, (Presse Universitaires de France), III-4-1.
- [18] G.H. Rees, "Injection" and "Extraction", CERN 85-19, (Nov. 1985), 331–45, and 346–57.
- [19] M.H. Blewett, "The Cosmotron - A Review", Rev. Sci. Instr., Vol. 24, No. 9, (Sept. 1953), 725-37.
- [20] K.R. Symon, A.M. Sessler, "Methods of radio frequency acceleration in fixed field accelerators with applications to high current intersecting beam accelerators", Proc. CERN Symposium on High Energy Accelerators and Pion Physics, (CERN 1956), 44–58.
- [21] G.I. Budker, G.I. Dimov, "On the charge exchange injection of protons into ring accelerators", Proc. 5th Int. Conf. on High Energy Accelerators, Dubna, 1963, (translation Conf-114, US, AEC, Div. of Tech. Info., 1372-7).
- [22] J. Liouville, "Sur la théorie de la variation des constantes arbitraires", J. de Math. Pures et Appl., 3, (1838), 342.
- [23] V.C. Kempson, C.W. Planner, V.T. Pugh, "Injection dynamics and multi-turn charge-exchange injection into the fast cycling synchrotron for the SNS", Proc. 1981 Part. Accel. Conf., Washington, IEEE Trans. Nucl. Sci., Vol NS-28, No. 3, (June 1981), 3085–7
- [24] L.C. Teng, "Design concepts for the beam scaper system of the main ring", Int. Report. Fermilab, FN-196/0400, (17 Oct. 1969).
- [25] J.B. Jeanneret, "Phase difference between collimators in a collider", SL/EA/Note 90-01.
- [26] The LHC Study Group, "Design study of the large hadron collider (LHC)", CERN 91-03, 115–25.
- [27] M. Seidel, "The HERA-p collimation system and first experience with a single collimator", HEACC 92, Hamburg, (to be published).
- [28] P.J. Bryant, E. Klein, "The design of betatron and momentum collimation systems". CERN SL/92-40, (August 1992).
- [29] T. Risselada, R. Jung, D. Neet, H. O'Hanlon, L. Vos, "The ISR collimator system", 1979 Part. Accel. Conf., San Francisco, NS-26, No. 3, 4131–3.
- [30] G. von Holtey, "Electron beam collimation at LEP energies", Proc. 1987 IEEE Part. Accel. Conf., Washington, (IEEE Cat. No. 87CH2387-9), Vol. 2, 1252-4.
- [31] P.J. Bryant, "Some advantages of inclined collimators", CERN SL/92-24 (AP).
- [32] T. Trenkler, "The optimum phase advance in a two-stage collimator system", CERN SL-92-50(EA).
- [33] R.B. Palmer, Proc. DPF Summer Study Snowmass 88, Snowmass, CO, 1988, SLAC-PUB-4707, (1988).
- [34] J. Bosser, "Electron cooling", Proc. CERN Accelerator School Fourth Advanced Course, Noordwijkerhout, 1991, CERN 92-01, (1992), 147–203.

# CHROMATICITY

*S. Guiducci*

Frascati National Laboratories - INFN, Frascati, Italy

## Abstract

A derivation of chromaticity formulae for circular accelerators is given, including the contribution of quadrupoles and sextupoles and with special attention to the general bending magnet. These are also exact for small radii of curvature.

## 1. INTRODUCTION

In the design of storage rings there are many similarities with the geometry of optics. In analogy to chromatic aberrations in optics, in particle accelerators a parameter called chromaticity is introduced. In optics rays of different wavelength find a different refraction index in a lens and therefore experience a different focal length. Similarly in a storage ring particles of different momentum see a different focusing strength in the quadrupoles and, as a consequence, have a different betatron oscillation frequency.

We define the chromaticity as the variation of the betatron tune  $Q$  with the relative momentum deviation  $\delta$  ( $\delta = \Delta p/p$ ):

$$Q' = \frac{dQ}{d\delta} . \quad (1)$$

Sometimes the relative chromaticity  $\xi$  is used:

$$\xi = \frac{Q'}{Q} . \quad (2)$$

Let us point out the importance of the chromaticity in circular accelerators. The chromaticity has a deleterious influence on the beam dynamics for two main reasons:

First, a momentum spread  $\sigma_p$  is always present in a particle beam, therefore the chromaticity produces a tune spread in the beam:

$$\Delta Q = Q' \sigma_p . \quad (3)$$

In large rings, with high tune values, this tune spread is so large that it is impossible to accommodate the beam in the space between the resonance lines in the tune diagram. How dangerous these resonances can be for beam stability has been described by E.J.N. Wilson [1].

Second, in the case of bunched beams the chromaticity produces a transverse instability called "head-tail effect" (see Ref. [2] for a detailed treatment). The wake field produced by the leading part of a bunch (the head) excites an oscillation of the trailing part (the tail) of the same bunch. In half a synchrotron period the head and the tail of the bunch interchange their positions and the oscillation can be anti-damped and may cause a beam loss. A complete mathematical treatment shows that the growth rate of this instability is much faster for negative than for positive chromaticity values and vanishes for zero chromaticity. It may be counteracted by a transverse feedback system, but this makes machine operation much more critical. Therefore most of the storage rings operate with zero or slightly positive chromaticity.

The "natural" chromaticity of a storage ring is that due only to the elements of the linear lattice, i.e. quadrupoles and dipoles. As will be shown later the "natural" chromaticity of a strong focusing storage ring is always negative and cannot be made zero. To correct the chromaticity nonlinear elements, the sextupole magnets, have to be introduced into the lattice.

In strong focusing lattices the main contribution to the chromaticity is due to the quadrupoles, in particular, in large rings with very large radius the contribution of the dipoles can be neglected; for small rings, however, the dipole contribution is important and has to be carefully calculated.

In Sections 2 and 3 it is shown how to calculate the chromaticity due to the quadrupoles and sextupoles respectively. Then, in Section 4, the effects on beam dynamics due to the chromaticity correcting sextupoles are briefly discussed. Finally, in Section 5, a detailed derivation of the chromaticity for a general bending magnet is given, following the approach given by M. Bassetti in Ref. [3], which is very simple and intuitive, avoiding long mathematical derivations.

## 2. QUADRUPOLE

Let us consider the motion in a quadrupole magnet of a charged particle which obeys the betatron equation:

$$y'' + k_y y = 0 \quad (y = x \text{ or } z) \quad (4)$$

with

$$k_x = -k$$

$$k_z = k .$$

Now we consider the dependence of  $k$  on the particle momentum  $p$ :

$$k = \frac{e}{p} \frac{\partial B_z}{\partial x} = \frac{e}{p_0(1+\delta)} \frac{\partial B_z}{\partial x} = \frac{k_0}{(1+\delta)} \simeq k_0 (1 - \delta + \delta^2 - \dots) \quad (5)$$

where  $\delta$  is the relative momentum deviation respect to the reference particle. Taking the first order term in  $\delta$  it is:

$$k \simeq k_0 (1 - \delta) = k_0 - k_0 \delta . \quad (6)$$

The chromatic variation has always the opposite sign with respect to the focusing strength, therefore a particle with a larger energy sees a weaker focusing strength. Conversely for light, the variation of the refraction index with the wavelength can be either positive or negative and the chromatic effect can be corrected to first order by combining lenses of different material.

Substituting (6) into the equation of motion for a quadrupole yields:

$$y'' = -k_y (1 - \delta) y \quad (7)$$

which is equivalent to adding to the focusing quadrupole a defocusing one with a strength  $-k_y \delta$  and viceversa for the defocusing quadrupole.

In a thin section of a quadrupole of infinitesimal length  $ds$  the particle receives an angular kick

$$dy' = y'' ds = k_y \delta y ds \quad (8)$$

described by a thin lens (defocusing for the focusing quadrupole) matrix:

$$\begin{pmatrix} 1 & 0 \\ k_y \delta ds & 1 \end{pmatrix} \quad (9)$$

To compute the effect of this kick on the betatron tune, the one-turn matrix is obtained by multiplying the unperturbed matrix ( $\delta = 0$ ) by this thin lens.

$$\mathbf{M} = \begin{pmatrix} \cos\mu_y & \beta_y \sin\mu_y \\ -\sin\mu_y/\beta_y & \cos\mu_y \end{pmatrix} \begin{pmatrix} 1 & 0 \\ k_y \delta ds & 1 \end{pmatrix} = \begin{pmatrix} \cos\mu_y + \beta_y \sin\mu_y k_y \delta ds & \beta_y \sin\mu_y \\ -\sin\mu_y/\beta_y + \cos\mu_y k_y \delta ds & \cos\mu_y \end{pmatrix} \quad (10)$$

Then we compute the trace of  $\mathbf{M}$  to get the new value of  $\mu_y$  ( $\mu_y = 2\pi Q_y$ ):

$$\frac{1}{2} \operatorname{Tr} \mathbf{M} = \cos(\mu_y + d\mu_y) = \cos \mu_y + \frac{1}{2} \beta_y \sin\mu_y k_y \delta ds \quad (11)$$

since:

$$d(\cos\mu_y) = \cos(\mu_y + d\mu_y) - \cos\mu_y = -\sin\mu_y d\mu_y \quad (12)$$

we get:

$$d\mu_y = -\frac{1}{2} \beta_y k_y \delta ds \quad (13)$$

or

$$dQ_y = -\frac{1}{4\pi} \beta_y k_y \delta ds. \quad (14)$$

Integrating over all the ring circumference  $L$ , we obtain the total chromaticity for the two planes, horizontal and vertical:

$$\frac{\partial Q_x}{\partial \delta} = -\frac{1}{4\pi} \int_0^L \beta_x(s) k_x(s) ds \quad (15)$$

$$\frac{\partial Q_z}{\partial \delta} = -\frac{1}{4\pi} \int_0^L \beta_z(s) k_z(s) ds. \quad (16)$$

*From these formulae we can see why, in a ring with strong focusing lattice, the chromaticity is always negative.* A quadrupole is a focusing lens in one plane, either horizontal or vertical, and defocusing in the other one. The strength  $k_{x,z}$  is positive in the focusing plane (negative in the defocusing plane) and the chromaticity has the opposite sign, therefore it is negative when the quadrupole is focusing. In a strong focusing lattice the  $\beta_{x,z}$  functions take the maximum values at the focusing quadrupoles, and the minimum at the defocusing ones, for each plane. Therefore the total chromaticity of a ring is dominated by the contribution of the focusing quadrupoles, negative in both planes.

### 3. SEXTUPOLE

Special magnets, the sextupoles, are inserted in the accelerator's lattice to correct the natural chromaticity produced by the focusing elements. In a sextupole a charged particle passing off-center receives a kick proportional to the square of its displacement from the center, i.e. a sextupole acts like a quadrupole with a focusing strength proportional to the displacement of the closed orbit from the sextupole center. This allows the chromaticity to be corrected because for off-momentum particles the closed orbit is displaced with respect to the reference one by a quantity  $D\delta$ , where  $D$  is the dispersion function and  $\delta$  the momentum deviation.

The field of a sextupole is given by :

$$B_x = g' xz \quad (17)$$

$$B_z = \frac{1}{2} g' (x^2 - z^2)$$

with

$$g' = \frac{\partial B_z^2}{\partial x^2} .$$

In Fig. 1 an example of the pole shape of a sextupole magnet is given.

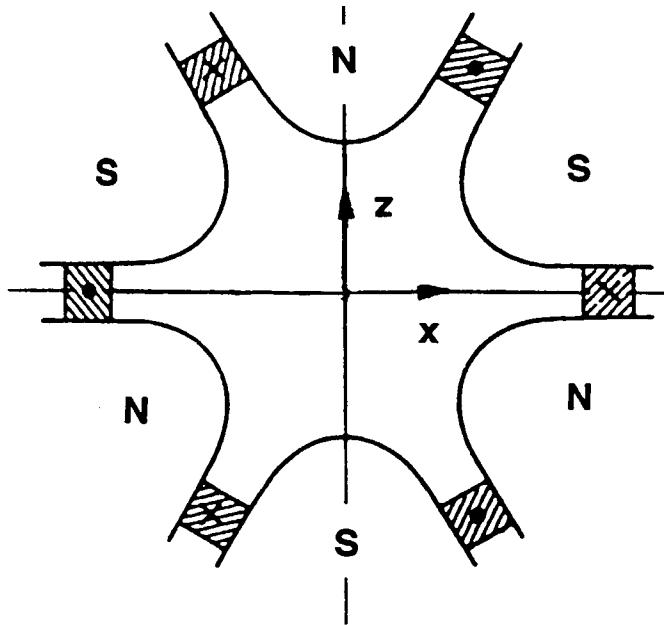


Fig. 1 Schematic representation of a sextupole magnet cross section

The equations of motion become:

$$x'' + \frac{1}{2} r (x^2 - z^2) = 0 \quad \text{where } r = \frac{e}{p_0} g' \quad (18)$$

$$z'' - rxz = 0 .$$

and the sextupole kick is :

$$\begin{aligned} dx' &= -\frac{1}{2} r (x^2 - z^2) ds \\ dz' &= r x z ds . \end{aligned} \quad (19)$$

Substituting the total coordinates for the off-momentum particle

$$x_t = D\delta + x \quad (20)$$

$$z_t = z$$

it becomes:

$$\begin{aligned} dx' &= - \left[ D\delta x + \frac{1}{2}(D\delta)^2 + \frac{1}{2}(x^2 - z^2) \right] r ds \\ dz' &= [D\delta z + xz] r ds . \end{aligned} \quad (21)$$

The first term of Eqs. (21) is equivalent to the kick of a quadrupole with gradient  $-rD\delta$  and, analogously to Eq. (14), gives a tune shift

$$\Delta Q = \frac{1}{4\pi} \beta r D \delta ds \quad (22)$$

and a contribution to the chromaticity:

$$\begin{aligned} \frac{\partial Q_x}{\partial \delta} &= \frac{1}{4\pi} \int_0^L \beta_x(s) r(s) D(s) ds \\ \frac{\partial Q_z}{\partial \delta} &= -\frac{1}{4\pi} \int_0^L \beta_z(s) r(s) D(s) ds . \end{aligned} \quad (23)$$

#### 4. CHROMATICITY CORRECTION

The most efficient way to correct chromaticity is to perform a localized correction, i.e. to insert a sextupole just in the same position of each quadrupole, where the chromatic effect is produced. In this ideal case the strength required to make the chromaticity zero is minimum and is simply related to the quadrupole strength:

$$r = -\frac{k}{D} \frac{l_Q}{l_S} \quad (24)$$

where  $l_Q$  and  $l_S$  are the lengths of the magnets. In practical cases one tries to put the chromaticity correcting sextupole as close as possible to each quadrupole but often, for economical reasons, less sextupoles with higher strengths are used.

In many cases, unfortunately, localized correction is not possible. For example collider storage rings have  $low-\beta$  insertions with very strong quadrupoles and zero dispersion function. Similarly, storage rings for synchrotron light production have many zero dispersion

straight sections for insertion devices, like wigglers and undulators, and strong focusing quadrupoles to get low emittances. In these cases a strong chromaticity produced in the insertions has to be corrected in the arcs and the sextupole strengths become very high.

If the arcs are built up by  $N$  periodic cells, two sextupoles are inserted in each cell, one in a high  $\beta_x$  place, to correct horizontal chromaticity, and the other in a high  $\beta_z$  position to correct the vertical one. The sextupole intensities are obtained by solving the following linear system of equations:

$$\begin{aligned} r_H l_S \beta_x^H D^H + r_V l_S \beta_x^V D^V &= \frac{Q'_x}{N} \\ -r_H l_S \beta_z^H D^H - r_V l_S \beta_z^V D^V &= \frac{Q'_z}{N}. \end{aligned} \quad (25)$$

where  $r_H$  and  $r_V$  are respectively the strengths of the horizontal and vertical chromaticity correcting sextupoles and  $Q'_x$ ,  $Q'_z$  are the values of the total chromaticity of the ring in the two planes. Since  $Q'_x$  and  $Q'_z$  are, usually, both negative it has to be:

$$r_H < 0 \quad \text{and} \quad r_V > 0$$

and each sextupole corrects the chromaticity in one plane and increases it in the other.

In order to reduce the sextupole strengths, it is important to place them where the dispersion is high and the  $\beta$  functions well separated. In fact if the horizontal sextupole is placed where  $\beta_x \gg \beta_z$  it is very effective in the horizontal plane and gives a negligible contribution in the other plane (viceversa for the vertical sextupole).

Generally the vertical sextupole has a strength higher than the horizontal one because the dispersion function  $D$  follows the behaviour of the horizontal  $\beta$ -function, i.e. it is higher at the horizontal sextupole and lower at the vertical one. This is specially true for collider storage rings, where, due to the low value of  $\beta_z$  at the interaction region, the vertical chromaticity is generally higher than the horizontal one.

The sextupoles necessary to correct the chromaticity introduce unwanted effects due to the other two terms in Eqs. (21):

- the chromatic aberration term  $(D\delta)^2$
- the geometric aberration term  $(x^2 - z^2), xz$ .

The geometric aberration term introduces higher-order terms in the equations of motion. In fact each sextupole inserted into the linear lattice, also in thin-lens approximation, doubles the order of the polynomial which links the initial and final coordinates for one turn in the ring. With  $N$  sextupoles in the ring the final coordinates depend on the  $2^N$ -th power of the initial one:

$$x(L) = a_{11} x(0) + a_{12} x'(0) + a_{13} \delta + \dots + a_{1j} x(0)^{2^N}.$$

When the nonlinear terms become important the stability of the particle trajectories in a circular accelerator is no longer obtained from the one-turn matrix  $\mathbf{M}$  ( $|\text{tr } \mathbf{M}| = |2 \cos \mu| < 2$ ), but depends on the amplitude of the betatron and synchrotron oscillations so that the beam dynamics becomes much more complicated.

In some very simplified cases an analytical calculation of the stability region is possible, for example in the unidimensional case ( $x, x'$  or  $z, z'$  phase plane) in the vicinity of a single resonance. In this case a closed curve, called the separatrix, can be found which divides the phase plane in two regions, a stable one inside the separatrix and an unstable region outside.

In more general cases tracking is used, i.e. a computer code which, given the initial coordinates for a particle in phase space, follows the evolution of a trajectory with the mathematical model chosen for the ring. A trajectory is considered stable if it remains confined in a certain phase space region for a given number of turns. The initial coordinates of the particle's tracking are changed to determine the largest region of phase space which contains all stable trajectories. This region is called the *dynamic aperture*.

This procedure is limited by computer time and precision, in fact the range of initial coordinates which can be explored in a six-dimensional phase space is very poor and the number of turns is always much smaller with respect to the beam lifetime or damping time.

After the linear lattice design a dynamic aperture optimization has to be carried out by choosing the distribution and the strengths of the sextupoles, the working point in the tune diagram, and even modifying the linear lattice to reduce chromaticities and sextupole strengths. A discussion of the methods for the determination and optimization of the dynamic aperture, a very important problem in the design of new accelerators, is given by A. Ropert [4].

## 5 . GENERAL BENDING MAGNET

In rings with large radius of curvature and small dispersion function the contribution to the chromaticity due to the bending magnets is negligible with respect to that of the quadrupoles. This is not the case for small rings. In the following is presented a derivation of the chromaticity formulae for the bending magnet valid also for small radii of curvature and taking into account the variation of the magnetic field in the ends.

In a bending magnet the betatron motion is given by the following equations:

$$y'' + k_y(s)y = 0 \quad (y = x \text{ or } z) \quad (27)$$

with

$$\begin{aligned} k_x &= -k + h^2 \\ k_z &= k \\ h &= -\frac{e}{p} B_z \quad ; \quad k = \frac{e}{p} \frac{\partial B_z}{\partial x} . \end{aligned}$$

The solution of these equations is represented, in each plane, by the two-by-two betatron matrix  $A$ . This matrix can be written as the product of  $N$  matrices  $A_i$ :

$$A = \prod_{i=1}^N A_i \left( \frac{l_B}{N} \right) \quad (28)$$

where  $l_B$  is the length of the bending magnet. We choose  $N$  to be large so that

$$\Delta s = \frac{l_B}{N} \rightarrow 0 .$$

This is equivalent to subdividing the magnet into  $N$  thin pieces of length  $\Delta s$ . To first

order in  $\Delta s$ ,  $A_i$  can be written as the product of a thin lens and a drift space:

$$A_i = \begin{pmatrix} 1 & 0 \\ -k_y(s)\Delta s & 1 \end{pmatrix} \begin{pmatrix} 1 & \Delta s \\ 0 & 1 \end{pmatrix} \quad (29)$$

Now we consider the changes that occur in the betatron motion (i.e. in the matrix  $A_i$ ) for a particle with a relative momentum deviation  $\delta$  oscillating around the off-momentum closed orbit.

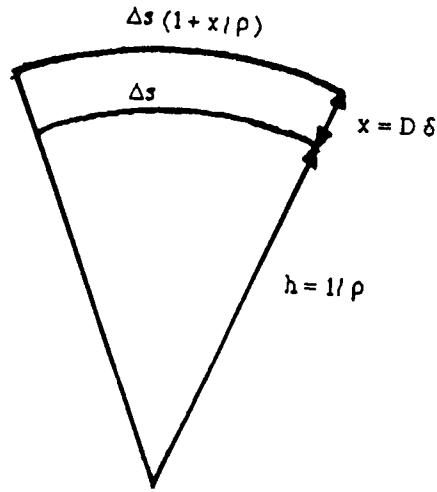


Fig. 2 Orbit lengthening

Two changes occur in the matrix  $A_i$ :

i) an orbit lengthening (see Fig. 2)

$$\Delta s(\delta) = \Delta s(1 + hD\delta) \quad (30)$$

ii) a change in the focusing strength of the thin lens due to:

- momentum dependence of the focusing functions
- variation of the length

$$- k_y(s, \delta) \Delta s(\delta) = - [k_y(s) + k_{Iy}(s)\delta + k_y(s)hD\delta] \Delta s \quad (31)$$

where  $k_{Iy}$  is the derivative:  $k_{Iy} = \partial k_y / \partial \delta$ .

As already seen, a change  $\Delta k$  in the focusing function at the position  $s$  gives a tune shift :

$$\Delta Q = - (1/4\pi) \beta(s) \Delta k \quad (32)$$

and, similarly, a change  $\Delta s$  in the length of a drift space gives:

$$\Delta Q = (1/4\pi) \gamma(s) \Delta s \quad (33)$$

where  $\gamma(s)$  is the Twiss function.

Integrating over all the circumference gives

$$\frac{\partial Q_y}{\partial \delta} = \frac{1}{4\pi} \int_0^L \{ \beta_y [ k_{Iy} + k_y h D ] + \gamma_y h D \} ds . \quad (34)$$

This formula is a generalization of that for a quadrupole, in fact for a quadrupole we have  $h = 0$  and  $k_{Iy} = dk_y/d\delta = -k_y$  and we obtain again the formulae of Eqs. (15) and (16).

In order to calculate  $k_{Iy}$  for the general bending magnet we need to know the fields seen by an off-momentum particle. First we write the second-order magnetic field expansion in the reference system of the design orbit for zero momentum deviation. The formulation of the field equations is that given by K. Steffen [5] with the only difference that  $h(s)$  has the opposite sign and its dependence on  $s$  is explicitly given, i.e:

$$h(s) = [h + h's + \frac{1}{2} h''s^2 + O(3)] . \quad (35)$$

As it will be useful in the following to distinguish the second-order terms they have been enclosed in square brackets:

$$\begin{aligned} B_z &= \frac{p}{e} \left\{ -h - h's + kx + \left[ \frac{1}{2} h''s^2 + k'xs + \frac{1}{2} rx^2 + \frac{1}{2} (h'' - hk - r) z^2 \right] + O(3) \right\} \\ B_x &= \frac{p}{e} \{ kz + [k'zs + rxz] + O(3) \} \\ B_s &= \frac{p}{e} \{ -h'z + [(hh' + k')xz] + O(3) \} . \end{aligned} \quad (36)$$

The previous equations are completely general, they are only based on the assumption of a field symmetry with respect to the median plane ( $z = 0$ ). Therefore, if we change the momentum of the particle, the origin and the orientation of the axis in the  $z = 0$  plane, the magnetic field has always the same form, but different values of the coefficients.

Now we make a transformation to the reference system of the off-momentum particle, as shown in Fig. 3:

$$\begin{aligned} p &= p^*/(1+\delta) \\ z &= z^* \\ x &= d + x^* \cos \theta + s^* \sin \theta \\ s &= -x^* \sin \theta + s^* \cos \theta \end{aligned} \quad (37)$$

where  $d = D\delta$  and  $\theta = D'\delta$ .

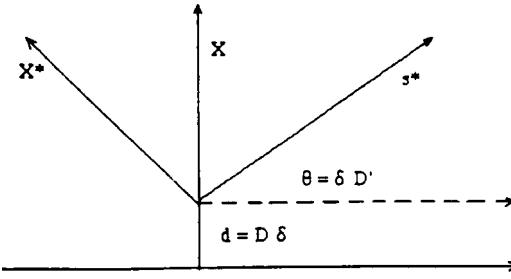


Fig. 3 Transformation of the reference system  
The field equations change in the following way:

$$\begin{aligned} B_z^* &= B_z[\vec{x}(\vec{x}^*)] \\ B_x^* &= B_x[\vec{x}(\vec{x}^*)] \cos\theta - B_s[\vec{x}(\vec{x}^*)] \sin\theta \\ B_s^* &= B_x[\vec{x}(\vec{x}^*)] \sin\theta + B_s[\vec{x}(\vec{x}^*)] \cos\theta. \end{aligned} \quad (38)$$

We are interested in the first-order field expansion, therefore we take only the first-order terms in Eqs. (36) and make the substitution :

$$\begin{aligned} B_z^* &= \frac{p}{e(1+\delta)} [-h + kd + x^*(k \cos\theta + h' \sin\theta) - s^*(h' \cos\theta - k \sin\theta)] \\ B_x^* &= \frac{p}{e(1+\delta)} z[k \cos\theta + h' \sin\theta] \\ B_s^* &= \frac{p}{e(1+\delta)} z[k \sin\theta - h' \cos\theta]. \end{aligned} \quad (39)$$

As already said, the various terms in the field equations have to be the same as in Eqs. (36), therefore equating the corresponding first-order terms we get the new coefficients:

$$\begin{aligned} h^* &= \frac{h-kd}{1+\delta} \\ k^* &= \frac{k \cos\theta + h' \sin\theta}{1 + \delta} \\ h'^* &= \frac{h' \cos\theta - k \sin\theta}{1 + \delta}. \end{aligned} \quad (40)$$

Using

$$\sin\theta \sim D'\delta; \quad \cos\theta \sim 1$$

and keeping only first-order terms in  $\delta$  we get:

$$\begin{aligned} k_x^* &= h^{*2} - k^* = (h^2 - k) + \delta(-2h^2 - 2hkD + k - h'D') \\ k_z^* &= k^* = k + \delta(-k + h'D'). \end{aligned} \quad (41)$$

We obtain the values of  $k_{Iy}(s)$  as:

$$k_{Ix} = \frac{\partial k_x^*}{\partial \delta} = k - 2h^2 - 2hkD - h'D' \quad (42)$$

$$k_{Iz} = \frac{\partial k_z^*}{\partial \delta} = -k + h'D'.$$

Inserting these values into Eq. (34) we obtain the final formulae:

$$\frac{\partial Q_x}{\partial \delta} = \frac{1}{4\pi} \int_0^L \{ \beta (k - 2h^2 - 2hkD - h'D') + \beta hD (h^2 - k) + \gamma hD \} ds \quad (43)$$

$$\frac{\partial Q_z}{\partial \delta} = \frac{1}{4\pi} \int_0^L \{ \beta (-k + hkD + h'D') + \gamma hD \} ds .$$

As we used only first-order terms in this derivation the contribution of the sextupole term  $\beta rD$ , calculated in Section 3, does not appear in Eqs. (43). In Appendix I a similar derivation using the second-order field expansion is given. The final formulae contain the same terms as Eqs. (43) plus the sextupolar terms coming from the second-order terms in the field expansion which are linear in  $x$  and, applying the translation  $x = x^* + D\delta$  of Eqs. (37), produce linear terms.

## 5. END-FIELD EFFECTS

From Eqs. (43) it is possible to calculate the contribution of the fringing fields to the chromaticity, once an expression for  $h'(s)$  is known. In a paper by W. Hardt, J. Jäger and D. Möhl [6] the same formulae are obtained with a different derivation, moreover a detailed calculation of the fringing field effects is given. For completeness we report the final formula obtained there. In Fig. 4 is reproduced the illustration of Ref. [6] which shows the schematization used for the end fields, the corresponding parameter definitions are listed below.

$s_1$	beginning of the central part
$s_2$	end of the central part
"1"	entrance of the fringing region
"2"	exit of the fringing region
$\theta$	entrance or exit angle of the trajectory
$\frac{1}{\tau \cos^3 \theta}$	radius of curvature of the end faces
$h = \frac{1}{\rho}$	curvature of the reference orbit
$k = -\frac{1}{B\rho} \frac{\partial B_z(0,0,s)}{\partial x}$	quadrupole component

$r = -\frac{1}{B\rho} \frac{\partial^2 B_z(0,0,s)}{\partial x^2}$	sextupole component
$D, D'$	dispersion function and its derivative
$\alpha, \beta, \gamma$	Twiss functions.

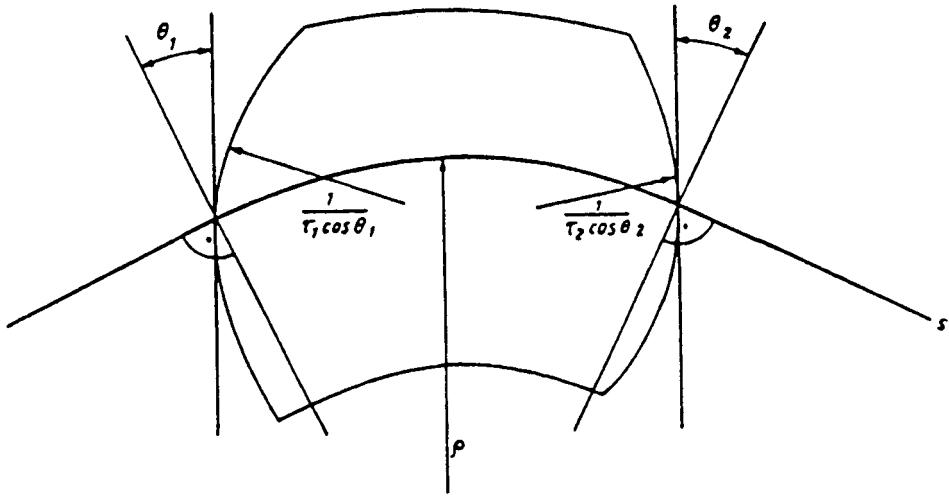


Fig. 4 Field boundaries for a bending magnet

The formulae to calculate the chromaticity of a magnet in terms of the lattice functions at the reference orbit are:

$$\begin{aligned} \frac{\partial Q_x}{\partial \delta} &= -\frac{1}{4\pi} \int_{s_1}^{s_2} [ (h^2 - k) \beta + rD\beta + h(2kD\beta + 2D'\alpha - D\gamma)] ds \\ &\quad + [-\operatorname{tg}\theta(h\beta + 2Dk\beta) + h\operatorname{tg}^2\theta(\beta D' - 2\alpha D + hD\beta\operatorname{tg}\theta) + \tau h\beta D] "1" \\ &\quad + [-\operatorname{tg}\theta(h\beta + 2Dk\beta) - h\operatorname{tg}^2\theta(\beta D' - 2\alpha D - hD\beta\operatorname{tg}\theta) + \tau h\beta D] "2" \end{aligned} \quad (44)$$

$$\begin{aligned} \frac{\partial Q_z}{\partial \delta} &= -\frac{1}{4\pi} \int_{s_1}^{s_2} [ k\beta - rD\beta - h(kD\beta + D\gamma)] ds \\ &\quad + [\operatorname{tg}\theta(h\beta + 2Dk\beta) - h\operatorname{tg}^2\theta(\beta D' - 2\alpha D - hD\beta\operatorname{tg}\theta) - \beta h D' - \tau h\beta D] "1" \\ &\quad + [\operatorname{tg}\theta(h\beta + 2Dk\beta) + h\operatorname{tg}^2\theta(\beta D' - 2\alpha D + hD\beta\operatorname{tg}\theta) + \beta h D' - \tau h\beta D] "2" \end{aligned}$$

## **ACKNOWLEDGEMENTS**

The calculation of the chromaticity for the bending magnet presented in this lecture is based on a paper written by M. Bassetti. I wish to thank him for the suggestions and helpful discussions during the preparation of this note.

## **REFERENCES**

- [1] E.J.N. Wilson, Non-linearities and resonances, these proceedings.
- [2] J.-L. Laclare, Introduction to coherent instabilities, *Ibid.*
- [3] M. Bassetti, A simplified derivation of chromaticity formulae, CERN LEP Note 504 (1984).
- [4] A. Ropert, Dynamic aperture, Proceedings CERN Accelerator School, Uppsala, Sept. 1989, CERN 90-04 (1990).
- [5] K. Steffen, High energy beam optics, *Wiley* (1965) and also Basic course on accelerator optics, CAS, Gif-sur-Yvette, Sept. 1984, CERN 85-19 (1985).
- [6] W. Hardt, J. Jäger, D. Möhl, A general analytical expression for the chromaticity of accelerator rings, CERN PS/LEA/82-5 (1982).

## APPENDIX 1

### CHROMATICITY FOR A BENDING MAGNET TAKING INTO ACCOUNT SECOND-ORDER TERMS

#### A1.1 DERIVATION OF THE CHROMATICITY FORMULAE

The chromaticity formulae in Section 5 are obtained using the first-order expansion of the magnetic field. To obtain an expression for the chromaticity which contains also the sextupolar terms, the same derivation is repeated here using also the second-order terms in the field expansion given by Eqs. (36). To get the new expressions for the magnetic field we apply the reference system transformation described in Section 5 to the second-order field expansion of Eqs. (36).

Following the derivation given in Section 5 and using :

$$\sin\theta \sim D'\delta \quad ; \quad \cos\theta \sim 1$$

we get the expressions for the magnetic field in the new reference system:

$$B_z^* = \frac{p}{e(1+\delta)} [-h -h'(-x^*D'\delta + s^*) + k(D\delta + x^* + s^*D'\delta) + k'(D\delta + x^* + s^*D'\delta)(-x^*D'\delta + s^*) + \frac{1}{2} r(D\delta + x^* + s^*D'\delta)^2 + \frac{1}{2}(h'' - hk - r) z^2]$$

$$B_x^* = \frac{p}{e(1+\delta)} z \{ k + k'(-x^*D'\delta + s^*) + r(D\delta + x^* + s^*D'\delta) - [-h' - h''(-x^*D'\delta + s^*) + (hh' + k')(D\delta + x^* + s^*D'\delta)] D'\delta \} \quad (45)$$

$$B_s^* = \frac{p}{e(1+\delta)} z \{ [k + k'(-x^*D'\delta + s^*) + r(D\delta + x^* + s^*D'\delta)] D'\delta - h' - h''(-x^*D'\delta + s^*) + (hh' + k')(D\delta + x^* + s^*D'\delta) \} .$$

Neglecting the second-order terms, except for the chromatic ones, i.e. the terms  $x\delta$ ,  $z\delta$  and  $s\delta$ , we obtain:

$$B_z^* = \frac{p}{e(1+\delta)} [-h + kD\delta + x^*(h'D'\delta + k + rD\delta) + s^*(-h' + kD'\delta + k'D\delta)]$$

$$B_x^* = \frac{p}{e(1+\delta)} z[k + rD\delta + h'D'\delta] \quad (46)$$

$$B_s^* = \frac{p}{e(1+\delta)} z[-h' + kD'\delta + (hh'+k')D\delta] .$$

Comparing these equations with Eqs. (36) and equating the corresponding first-order terms, we get the new coefficients:

$$\begin{aligned} h^* &= \frac{h-kD\delta}{1+\delta} \\ k^* &= \frac{k + h'D'\delta + rD\delta}{1 + \delta} \\ h'^* &= \frac{h' + kD'\delta + (hh'+k')D\delta}{1 + \delta} . \end{aligned} \quad (47)$$

Now, following the same procedure as in Section 5, we use the coefficients  $h^*$  and  $k^*$  given by Eqs. (47) to obtain the values of the focusing strength for the off-momentum particle:

$$\begin{aligned} k_x^* &= h^{*2} - k^* = (h^2 - k) + \delta(-2h^2 - 2hkD + k - h'D' - rD) \\ k_z^* &= k^* = k + \delta(-k + h'D' + rD). \end{aligned} \quad (48)$$

Then, we get the variation of the focusing strength with momentum,  $k_{Iy}(s)$ :

$$\begin{aligned} k_{Ix} &= \frac{\partial k_x^*}{\partial \delta} = k - 2h^2 - 2hkD - h'D' - rD \\ k_{Iz} &= \frac{\partial k_z^*}{\partial \delta} = -k + h'D' + rD . \end{aligned} \quad (49)$$

The variation of the orbit length with momentum has been already taken into account in Eq. (34), therefore inserting Eqs. (49) into (34) we obtain the final formulae for the chromaticity, which are more complete than those of Eqs. (43) because they contain also the sextupolar terms.

$$\begin{aligned} \frac{\partial Q_x}{\partial \delta} &= \frac{1}{4\pi} \int_0^L \{ \beta(k - 2h^2 - 2hkD - h'D' - rD) + \beta h D (h^2 - k) + \gamma h D \} ds \\ \frac{\partial Q_z}{\partial \delta} &= \frac{1}{4\pi} \int_0^L \{ \beta(-k + hkD + h'D' + rD) + \gamma h D \} ds . \end{aligned} \quad (50)$$

### A1.1.1 An observation on Eqs. (47)

Let us notice that the coefficient  $h'^*$  given by Eqs. (47) is obtained as the coefficient of the variable  $z$  in the equation (46) for  $B_s$ , and that it is different from the coefficient of  $s$  which appears in the expression for  $B_z$ . This ambiguity comes from the fact that, while for Eqs. (36) the relation:

$$\frac{\partial B_z}{\partial s} = \frac{\partial B_s}{\partial z} \quad (51)$$

is valid, this is not true for Eqs. (46), for which it is:

$$\frac{\partial B_s^*}{\partial z} \neq \frac{\partial B_z^*}{\partial s^*} . \quad (52)$$

Equations (46) are anyway correct, but the new variable  $s^*$  has to be modified. In cylindrical coordinates  $(z, x, \phi)$ , the radial component of the Maxwell equation is written:

$$\frac{1}{\rho} \frac{\partial B_z}{\partial \phi} = \frac{\partial B_\phi}{\partial z} . \quad (53)$$

When making the transformation given by Eqs. (37), which is essentially a translation in the radial direction, in Eqs. (53)  $\rho$  has to be replaced by  $\rho + D\delta$ . As a consequence, the Maxwell equation is written:

$$\frac{1}{\rho+D\delta} \frac{\partial B_z^*}{\partial \phi} = \frac{1}{1+hD\delta} \frac{\partial B_z^*}{\partial s^*} = f(\partial B_s^*, \partial z) ; \quad h = \frac{1}{\rho} . \quad (54)$$

This relation is in effect verified by Eqs. (46) to first order in  $\delta$ .

# A SIMPLE THEORY FOR WEAK BETATRON COUPLING

P.J. Bryant

CERN, Geneva, Switzerland

## ABSTRACT

The model of a sinusoidally smooth machine with uniform weak coupling fields is used to develop a simple theory for betatron coupling on the second order resonance  $Q_x - Q_z = 0$ . A coupling coefficient is defined. Observations and measurement methods are discussed, which despite their simple derivation are still valid. The normal modes are found in skew quadrupole and axial fields. A more exact interpretation of the coupling coefficient, which can be derived using the Hamiltonian formalism is quoted and briefly discussed. Knowledge of this detailed expression is required for the design of coupling compensation schemes.

## 1. INTRODUCTION

The periodic energy exchange between two oscillators, or coupling, is a widespread phenomenon in physics. The coupling action can take place via the amplitudes, velocities or accelerations of the oscillators. A classic example of amplitude coupling is a pair of pendulums linked by a weightless spring. This example also shows rather clearly the normal modes of the system (see Fig. 1). Any motion of the system can be expressed as a sum of these modes.

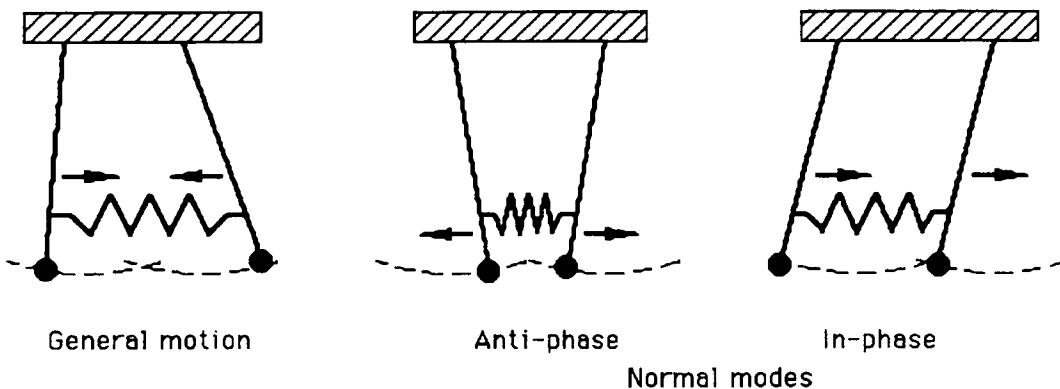


Fig. 1 Amplitude coupled pendulums

The first storage ring projects in the 1960's and early 1970's stimulated an interest in betatron coupling. Initially the concern was over the coupling from random quadrupole tilts, which introduce skew quadrupole field errors around the ring, and later over the axial fields of solenoids used as physics detectors. References 1 and 2 are early examples of the analysis of skew quadrupole errors in electron storage rings. There are two rather comprehensive references in the subsequent literature, Ripkin [3] in which the Courant, Livingston and Snyder theory is extended to a 4-dimensional phase space, and Guiqnard [4] in which the theory of all sum and difference resonances in 3-dimensional magnetic fields is developed by a perturbative treatment of the Hamiltonian for the single particle motion in a synchrotron. The method and essential results of this last reference were in fact anticipated in an addendum (July 1966) written for an internal SLAC note SR-11 (June 1966) by Morton [5]. The present paper develops a simple theory for betatron coupling in skew quadrupole and axial fields based on work done for the ISR in the 1970's [6-8]. The results from the more detailed theory [4] are simply quoted when needed.

## 2. COUPLING IN UNIFORM SKEW QUADRUPOLE AND AXIAL FIELDS

### 2.1 Basic equations

Coupling is a phenomenon which takes place over very many turns, so that it seems likely that average parameters such as the tunes are of more importance than the local lattice functions. On this premise, we will replace the equations of motion in the horizontal and vertical planes by sinusoidal approximations. We will also make the simplifying assumption that the coupling fields are spread uniformly around the ring. This can be regarded as considering the zero order harmonic of the distribution.

There are two categories of coupling fields:

- the transverse skew multipoles, which arise as errors from magnet tilts, and
- axial fields, which are usually introduced by physics detectors in colliders.

The first category causes amplitude coupling and the second velocity coupling. In accelerators, there is no equivalent of acceleration coupling as for example in inductancecoupled ac circuits,  $q_1 + LR_1/L_1|q_1 + L_1/L_1C_1J_1q_1 = -M_12q_2$

Equations (1) give the sinusoidal approximation to the betatron motion of a single particle in the presence of skew quadrupole and axial fields, which are uniform in distance and time.

$$\begin{aligned} x'' &= \kappa_x^2 x = -bz' - kz \\ z'' &= \kappa_z^2 z = bx' - kx \end{aligned} \quad (1)$$

where

$$\kappa_x = \frac{Q_x}{R} \quad \text{and} \quad \kappa_z = \frac{Q_z}{R} \quad (2)$$

R is the average machine radius and  $Q_x$  and  $Q_z$  the horizontal and vertical tunes respectively.

$$k = \frac{1}{B\rho} \left( \frac{\partial B_x}{\partial x} \right)_0 = \frac{-1}{B\rho} \left( \frac{\partial B_z}{\partial z} \right)_0 \quad \text{and} \quad b = \frac{1}{B\rho} B_{\text{axial}} \quad (3)$$

$B\rho$  is the magnetic rigidity,

' denotes d/ds differentiation along the beam path.

Let us try solutions to (1) which have the form of the uncoupled oscillations modified by an envelope term i.e.  $x = X(s) e^{ik_x s}$  and  $z = Z(S) e^{ik_z s}$ . By restricting ourselves to weak coupling conditions, we can neglect the terms in  $X''$  and  $Z''$ , since X and Z will be slowly varying functions. Hence,

$$2i\kappa_x X' = e^{-i\delta s} [-bZ' - (k + i\kappa_z b)Z] \quad (4)$$

$$2i\kappa_z Z' = e^{i\delta s} [bX' - (k - i\kappa_x b)X] \quad (5)$$

where,

$$\delta = (\kappa_x - \kappa_z) = (Q_x - Q_z)/R . \quad (6)$$

By substituting for  $Z'$  in (4) using (5) and vice versa, it can be seen that the  $X'$  and  $Z'$  terms on the right hand side are second order. By neglecting the  $b^2$  and  $bk$  terms, we get,

$$\begin{aligned} X' &= i \frac{e^{-i\delta s}}{2\kappa_x} (k + i\kappa_z b) Z \\ Z' &= i \frac{e^{i\delta s}}{2\kappa_z} (k - i\kappa_x b) X . \end{aligned} \quad (7)$$

Since we are interested in the second order coupling resonance  $Q_x \sim Q_z$  and  $\kappa_x = \kappa_z$ . Thus, in Eqs. (7) we put  $\kappa = \kappa_x = \kappa_z$  and maintain  $\delta = (\kappa_x = \kappa_z)$  to get,

$$X' = \frac{ie^{-i\delta s}}{2\kappa} (k + i\kappa b) Z \quad (8)$$

$$Z' = \frac{ie^{i\delta s}}{2\kappa} (k - i\kappa b) X . \quad (9)$$

Equations (8) and (9) are now in the standard form for coupled mode analysis (i.e.  $X' = c_{12}Z$ ,  $Z' = -c_{12}^*X$ , where the minus sign indicates positive energy flow in both oscillations and  $*$  indicates the complex conjugate).

In obtaining (8) and (9), we put  $\kappa = \kappa_x = \kappa_z$ . This approximation is essential to keep the coupling coefficients of equal magnitude to ensure the equations are conservative. Efforts to make the above analysis more realistic often result in violating this requirement. The use of the Hamiltonian formalism and conjugate variables avoids this pitfall and gives a rigorous result for a practical lattice (see Section 4).

Equations (8) and (9) can be solved to give the general solutions:

$$x = e^{i\kappa_x s} \left[ A e^{(i/2)(\eta-\delta)s} - B \frac{(\eta-\delta)}{k - i\kappa b} e^{-(i/2)(\eta+\delta)s} \right] \quad (10)$$

$$z = e^{i\kappa_z s} \left[ B e^{-(i/2)(\eta-\delta)s} + A \frac{(\eta-\delta)\kappa}{k + i\kappa b} e^{(i/2)(\eta+\delta)s} \right] \quad (11)$$

where

$$\eta = +\sqrt{\delta^2 + k^2/\kappa^2 + b^2} \quad (12)$$

A and B are complex constants

At first sight, Eqs. (10) and (11) do not convey a great deal. A greater physical insight is obtained by taking the squares of the moduli of the amplitude terms X and Z, which form the envelopes of the oscillations.

$$XX^* = |X|^2 = |A|^2 + \frac{|B|^2(\eta-\delta)^2}{(k^2/\kappa^2 + b^2)} - \frac{2|AB^*|(\eta-\delta)}{\sqrt{k^2/\kappa^2 + b^2}} \cos(\eta s - \phi) \quad (13)$$

$$ZZ^* = |Z|^2 = |B|^2 + \frac{|A|^2(\eta-\delta)^2}{(k^2/\kappa^2 + b^2)} - \frac{2|AB^*|(\eta-\delta)}{\sqrt{k^2/\kappa^2 + b^2}} \cos(\eta s - \phi) \quad (14)$$

where

$$\phi = \tan^{-1}[bk/\kappa] .$$

Thus there is a sinusoidal exchange of energy with the interchange wavelength,

$$\lambda = \frac{2\pi}{\eta} = \frac{2\pi}{+\sqrt{\delta^2 + k^2/\kappa^2 + b^2}} \quad (15)$$

and the sum of the squares of the amplitudes is conserved,

$$E_{max} = |X|^2 + |Z|^2 = \frac{2\eta}{(\eta + \delta)} (|A|^2 + |B|^2) . \quad (16)$$

It is worth considering the special case of a beam being kicked in one plane. The ensuing oscillations will be coherent and the amplitude interchanqe will be visible to a pick-up in either plane. Figure 2 illustrates this effect.

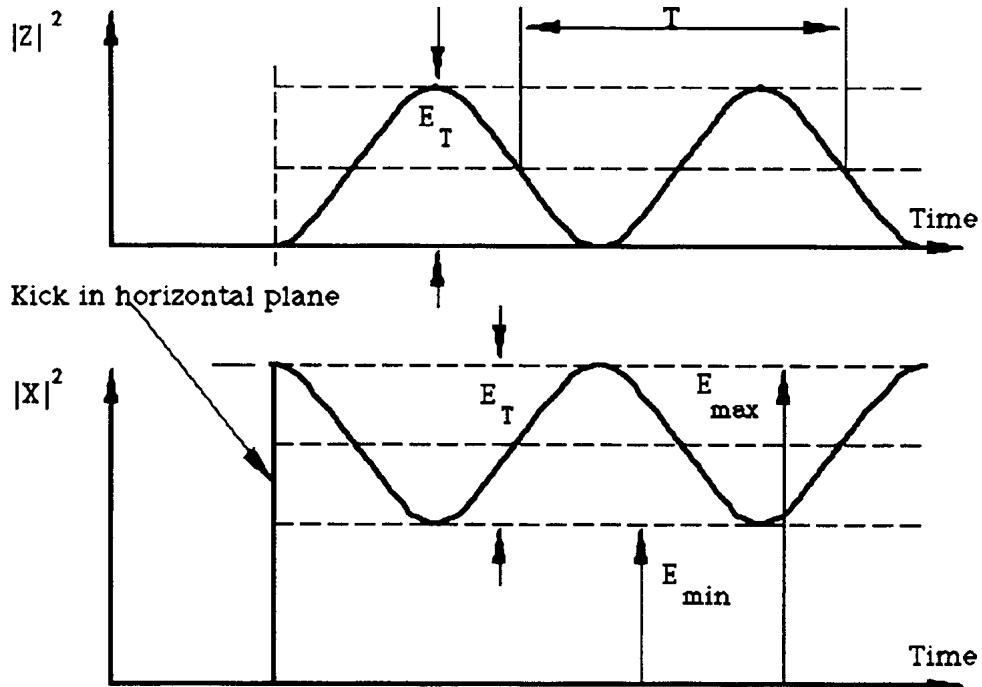


Fig. 2 Behaviour of the envelope functions of the coherent oscillations following a kick in one plane

In the case of Fig. 2, let  $s = 0$  at the kick, then  $|X|^2$  is a maximum ( $E_{max}$ ) and  $Z$  is zero (real and imaginary parts).

Equation (11) gives:

$$B = -\frac{A(\eta - \delta)\kappa}{(k + i\kappa b)}$$

and Eq. (13) gives the interchanged amplitude:

$$E_T = \frac{4|AB^*|(\eta - \delta)}{\sqrt{k^2/\kappa^2 + b^2}} .$$

Making use of the definition of  $\eta$  in (12), a little manipulation gives the modulation,

$$S = \frac{E_{min}}{E_{max}} = \frac{\delta^2}{\eta^2} . \quad (17)$$

At this stage it is convenient to define coupling coefficients, which effectively replace Eqs. (3),

$$\text{- for skew gradients, } C_q = \left( \frac{R^2}{Q} \right) \left( \frac{1}{B\rho} \right) \left( \frac{\partial B_x}{\partial x} \right)_0 \quad (18)$$

$$\text{- for axial fields, } C_b = \left( \frac{R}{B\rho} \right) B_{\text{axial}} . \quad (19)$$

These combine to give

$$|C| = +\sqrt{C_q^2 + C_b^2} . \quad (20)$$

Equations (15) and (17) can then be rewritten in terms of more easily observable quantities:

$$\text{Interchange period, } T = \frac{1}{f_{rev} \sqrt{\Delta^2 + |C|^2}} \quad (21)$$

$$\text{and Modulation, } S = \frac{E_{min}}{E_{max}} = \frac{\Delta^2}{\Delta^2 + |C|^2} \quad (22)$$

where

$$\begin{aligned} \Delta &= (Q_x - Q_z) \text{ (separation of uncoupled tunes)} \\ Q &= (Q_x + Q_z)/2 \text{ (average of uncoupled tunes)} \\ f_{rev} &= \text{revolution frequency} . \end{aligned} \quad (23)$$

Equations (21) and (22) give direct methods for measuring  $|C|$  by kicking a beam and observing the coherent oscillations. It should be stressed that  $\Delta$  is the separation of the uncoupled tune values, which cannot be measured directly in the presence of coupling, but by eliminating  $\Delta^2$  between (21) and (22) this problem is avoided.

$$|C| = \frac{1}{f_{rev} T} \sqrt{1 - S} \quad (24)$$

and

$$|\Delta| = \frac{\sqrt{S}}{f_{rev} T} . \quad (25)$$

This provides a very precise and elegant method for measuring  $|C|$ .

Despite the very simplified model used above, it turns out that (24) and (25) are still valid in the complete theory [4], which will be discussed in Section 4. The difference lies in how  $C_q$  and  $C_b$  should be interpreted in the case of a strong focusing lattice. For the moment, the first approximation to a real machine for  $C_q$  and  $C_b$  would be to replace (18) and (19) by:

$$C_q = \left( \frac{R}{2\pi Q} \right) \left( \frac{1}{B\rho} \right) \oint \left( \frac{\partial B_x}{\partial x} \right)_0 ds$$

$$C = \left( \frac{1}{2\pi} \right) \left( \frac{1}{B\rho} \right) B_{\text{axial}} ds .$$

## 2.2 Normal modes of oscillation

The normal modes are of interest because they give some physical insight into what is happening, which will enable us to find a method for measuring the two parts ( $C_q$  and  $C_b$ ) of the coupling coefficient.

A quick indication of the normal modes can be found by rewriting (10) and (11) to give:

$$x = Ae^{i(\kappa+\eta/2)s} - \frac{B(\eta-\delta)}{\sqrt{k^2/\kappa^2 + b^2}} e^{i(\kappa-\eta/2)s+i\phi} \quad (26)$$

$$z = Be^{i(\kappa-\eta/2)s} + \frac{A(\eta-\delta)}{\sqrt{k^2/\kappa^2 + b^2}} e^{i(\kappa+\eta/2)s-i\phi} \quad (27)$$

using

$$\kappa_x = \left( \kappa + \frac{\delta}{2} \right); \quad \kappa_z = \left( \kappa - \frac{\delta}{2} \right); \quad (28)$$

and where:

$$\phi = \tan^{-1}\left(\frac{\kappa b}{k}\right) = \tan^{-1}\left(\frac{C_b}{C_q}\right). \quad (29)$$

The normal modes can now be found by putting either of the two arbitrary constants A and B to zero. This gives:

$$\frac{x}{z} = \left( \frac{\Delta \pm \sqrt{\Delta^2 + |C|^2}}{C} \right) e^{i\phi}. \quad (30)$$

Equation (30) shows the characteristic of a normal mode that x and z bear a constant relationship i.e. the mode retains its "shape" at all times. The interpretation of (30) is that the normal modes lie on inclined planes and the complex phase shift indicates that they are elliptically polarised (see Figs. 3 and 4). This information helps us to solve the basic equations (1) and (2) in a different way. For a more detailed treatment of the normal modes see reference 6.

We rewrite (1) and (2) using (28) to get,

$$x'' + (\kappa^2 + \delta\kappa)x = -bz' - kz \quad (31)$$

$$z'' + (\kappa^2 - \delta\kappa)z = bx' - kx \quad (32)$$

where the terms  $\delta^2/4$  have been neglected on the left hand side.

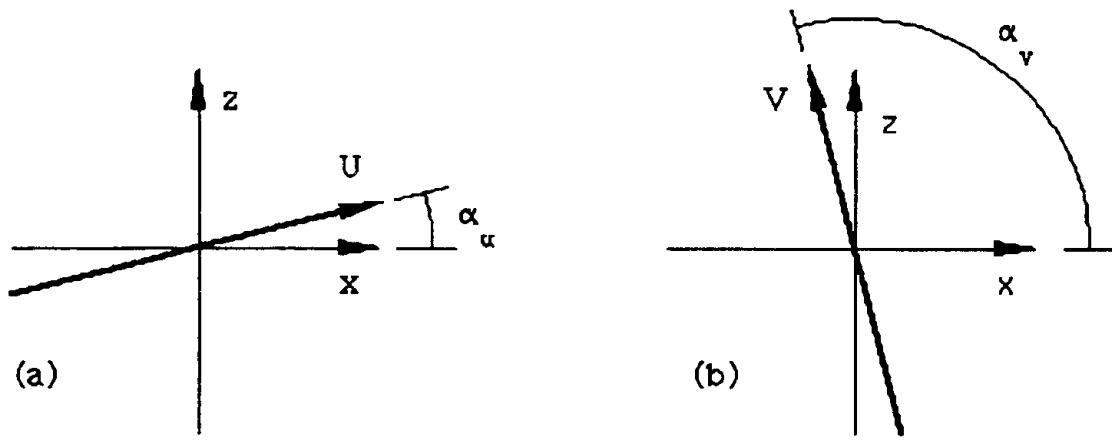


Fig. 3 Inclined (U,V) normal modes with skew quadrupole coupling. U – "horizontal-like" mode, V – "vertical-like" mode.

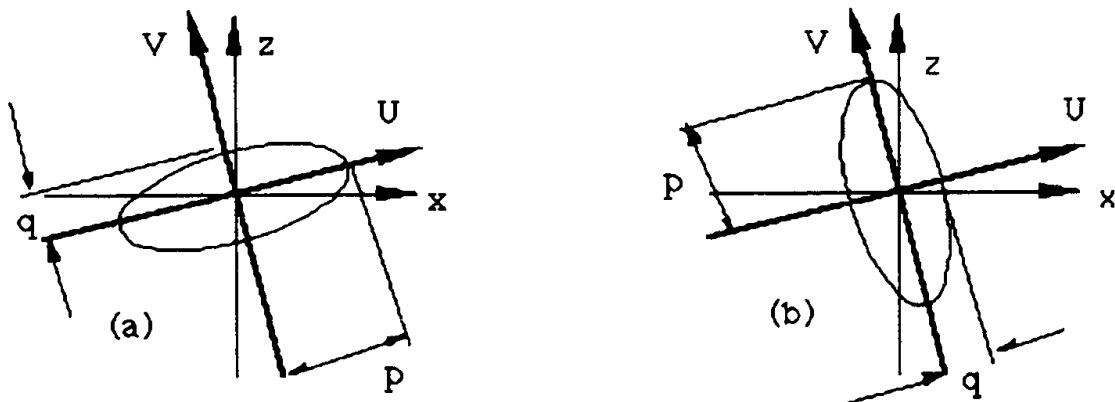


Fig. 4 Inclined elliptically polarised normal modes with skew quadrupole and axial field coupling

Equation (30) suggests that it will be possible to partially decouple (31) and (32) by making a transformation to a coordinate system (U,V), which is inclined at an angle  $\alpha$  to the (x,z) system, so that,

$$\begin{aligned} x &= U(s) \cos \alpha - V(s) \sin \alpha \\ z &= U(s) \sin \alpha + V(s) \cos \alpha . \end{aligned} \quad (33)$$

Substituting (33) into (31) and (32) yields two equations from which first  $V''$  can be eliminated and then  $U''$ , to give,

$$\begin{aligned} U'' \left( \tan \alpha + \frac{1}{\tan \alpha} \right) + U \left[ (\kappa^2 - \delta \kappa) \tan \alpha + \frac{(\kappa^2 + \delta \kappa)}{\tan \alpha} + 2k \right] \\ = -V' b \left( \tan \alpha + \frac{1}{\tan \alpha} \right) + V \left[ k \left( \tan \alpha - \frac{1}{\tan \alpha} \right) + 2\delta k \right] \end{aligned} \quad (34)$$

$$\begin{aligned}
& V'' \left( \tan \alpha + \frac{1}{\tan \alpha} \right) + V \left[ (\kappa^2 + \delta\kappa) \tan \alpha + \frac{(\kappa^2 - \delta\kappa)}{\tan \alpha} - 2k \right] \\
& = U' b \left( \tan \alpha + \frac{1}{\tan \alpha} \right) + U \left[ k \left( \tan \alpha - \frac{1}{\tan \alpha} \right) + 2\delta k \right]
\end{aligned} \tag{35}$$

By choosing  $\alpha$  such that the last terms in (34) and (35) are zero, the equations become decoupled for  $k$ . This value of  $\alpha$  is given by:

$$\tan \alpha = \frac{-\delta \pm \sqrt{\delta^2 + k^2 / \kappa^2}}{k / \beta} = \frac{-\Delta \pm \sqrt{\Delta^2 + C_q^2}}{C_q}. \tag{36}$$

This angle of inclination can be identified in (30) by putting  $C_b = 0$ . The two solutions can be called  $\alpha_u$  and  $\alpha_v$  and it is quickly verified that  $(\alpha_u - \alpha_v) = \pi/2$ .

Equation (36) yields the skew quadrupole coefficient directly as,

$$C_q = \Delta \tan 2\alpha. \tag{37}$$

Thus at the privileged angle of  $\alpha$

$$\begin{aligned}
U'' + (\kappa^2 + \gamma\kappa)U &= -bV' \\
V'' + (\kappa^2 - \gamma\kappa)V &= bU'
\end{aligned} \tag{38}$$

where

$$\gamma\kappa = \delta\kappa \cos 2\alpha + k \sin 2\alpha. \tag{39}$$

In the rotated U-V system only the axial field coupling is apparent, except for the shift  $\pm\gamma/2$  to get the "uncoupled" frequencies of the U-V system ( $k \pm \gamma/2$ ). Thus from (39), using (2), (3), (6), (19) and (23)

$$\Delta_{UV} = (\Delta \cos 2\alpha + C_q \sin 2\alpha).$$

It is convenient to eliminate  $C_q$  by using (37)

$$\Delta_{UV} = \Delta / \cos 2\alpha. \tag{40}$$

We can now apply the Eqs. (21) and (22) of Section 2 directly to the U-V system to get,

$$T_{UV} = \frac{1}{f_{rev} \sqrt{\Delta_{UV}^2 + C_b^2}} \tag{41}$$

$$S_{UV} = \frac{\Delta_{UV}^2}{\Delta_{UV}^2 + C_b^2}. \tag{42}$$

By substituting (37) and (40) in (41), it can be shown that  $T_{UV} = T$  of the original x-z system. Thus the period of the amplitude exchange is unchanged and only  $S$  and  $\sim$  change between the two systems. If we now eliminate  $\Delta_{UV}$ , as before, we find,

$$C_b = \frac{1}{f_{rev} T} \sqrt{1 - S_{UV}}. \tag{43}$$

If therefore, we could mount a kicker and pickup at variable angle, then at the point that  $|C|$  fell to a minimum, we would have found the  $\alpha$  of (36), the coupling coefficient would be  $C_b$  from (43) and as a check  $C_q$  would be given by (37) using  $\sim$  from the original measurement with  $\alpha = 0$ .

### 3. OBSERVATIONS

#### 3.1 Tune measurements

In (26) and (27) there are two frequencies, which are the frequencies of the normal modes

$$\kappa_{\frac{u}{v}} = \frac{1}{R} \left[ Q \pm \frac{1}{2} \sqrt{\Delta^2 + C_q^2 + C_b^2} \right].$$

If  $C_q = C_b = 0$ , then  $\kappa_u = Q_x/R$  and  $\kappa_v = Q_z/R$ . When  $C_q \neq 0$  and  $C_b \neq 0$  their effect depends on their magnitude relative to  $\Delta$ . If  $\Delta$  is much larger (i.e. the working point is far from the  $Q_x - Q_z = 0$  resonance) then  $C_q$  and  $C_b$  have little impact on the value of the square root. Their effect then increases until at  $\Delta = 0$ , they are totally responsible for the split in the frequencies. However, we know that not only do the normal mode frequencies change, but also their inclination and shape. When  $\Delta$  is large relative to  $C_q$  and  $C_b$  the modes appear to be the usual horizontal and vertical oscillations with frequencies  $Q_x/R$  and  $Q_z/R$ . As  $C_q$  and  $C_b$  increase the modes rotate, until at  $\Delta = 0$ ,  $\alpha = 45^\circ$  from (36). In this situation the working point is exactly on the resonance  $Q_x - Q_z = 0$  and the machine is said to be fully coupled.

If one were to do this experimentally by increasing  $Q_x$  and decreasing  $Q_z$  in the region of the  $Q_x - Q_z = 0$  resonance, while trying to measure  $Q_x$  and  $Q_z$  using horizontal and vertical kickers with the corresponding pickups, then the results would be similar to Fig. 5.

The pickups see the frequencies of the U- and V-modes. While these modes are "nearly" horizontal and "nearly" vertical the measurements are reliable. As the modes rotate the horizontal pickup will start to "see" the V-mode as well as the U-mode and vice versa. Operational difficulties for the tune measurement will appear at around  $|C/\Delta| = 1.0$ . When  $|C/\Delta| > 1$  tune measurements will become unreliable and the tune readings will start to jump back and forth from the U-mode to the V-mode values, spanning the stopband in Fig. 5. At  $\Delta = 0$ , the pickups respond equally well to both modes, since they are now at  $45^\circ$  inclination. The difference in the mode frequencies ( $Q_u - Q_v$ ) equals  $|C|$ , which provides another way of measuring  $|C|$ . As  $|\Delta|$  increases on the far side of the resonance normality will be restored.

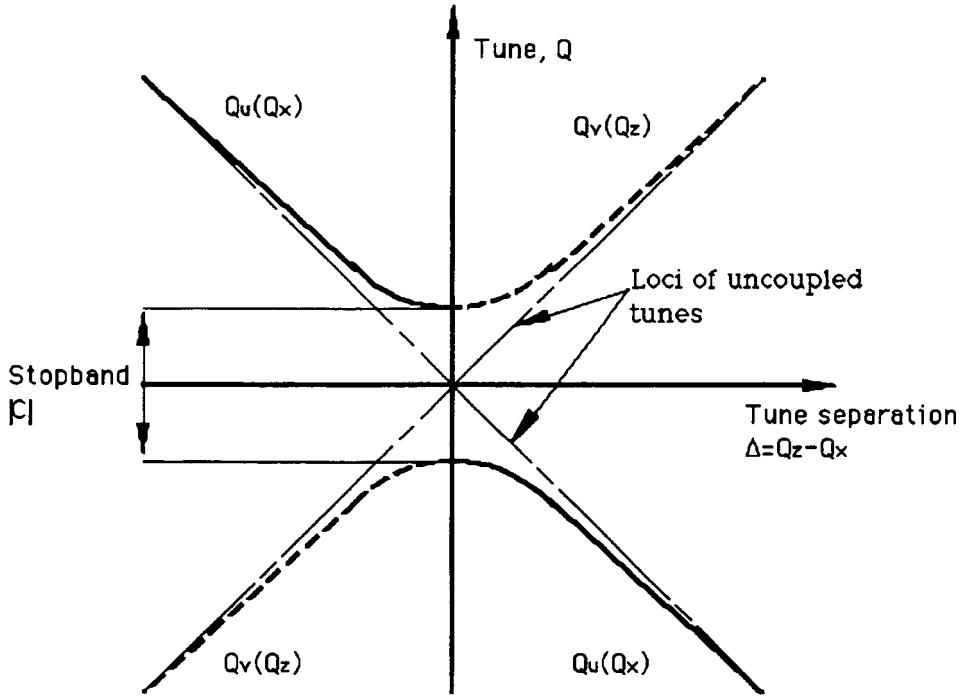


Fig. 5 Effect of coupling on tune measurements

### 3.2 Other measurement methods

We have already seen how to measure  $|C|$ ,  $|C_q|$  and  $|C_b|$  using a rotating kicker and pickup. The tune measurements in the previous section also gave a value for  $|C|$ . More information, such as the signs of  $C_q$  and  $C_b$  can be found in theory by carefully studying the waveforms following a kick [8], but this has never been demonstrated practically. A rather different method, which can be used on "quiet" beams, i.e. on physics beams in colliders, uses the "coupling transfer function", which is obtained by correlating a horizontal transverse perturbation imposed on a beam to the resulting vertical coherent motion (see Ref. 9).

## 4. RESULTS OF THE EXACT ANALYSIS [4]

The exact analysis for coupling in an alternating gradient lattice can be found in Ref. 4. This analysis confirms the formulae for  $T$ ,  $S$ ,  $\kappa_u$  and  $\kappa_v$ , but more importantly, it shows how to evaluate  $C$  for known errors in a lattice. In the formalism of Ref. 4,  $C$  is complex and is given by:

$$C(\theta_0) = \frac{1}{2\pi R} \int_{\theta_0}^{2\pi + \theta_0} \sqrt{\beta_x \beta_z} \left[ K(\theta) + \frac{M(\theta)}{2} R \left( \frac{\alpha_x}{\beta_x} - \frac{\alpha_z}{\beta_z} \right) - i \frac{M(\theta)}{2} R \left( \frac{1}{\beta_x} + \frac{1}{\beta_z} \right) \right] \\ \exp \left( i \left[ (\mu_x - \mu_z) - \{(\mu_x(\theta_0) - \mu_z)(\theta_0)\} - \Delta(\theta - \theta_0) \right] \right) d\theta \quad (44)$$

where  $\theta_0$  is origin at which  $C$  is being observed (evaluated) and  $\theta = s/R$

$$K(\theta) = \frac{1}{2} \frac{R^2}{|B\rho|} \left( \frac{\partial B_x}{\partial x} - \frac{\partial B_z}{\partial z} \right) \quad (45)$$

$$M(\theta) = \frac{R}{|B\rho|} B_{axial} . \quad (46)$$

The real part of  $C(\theta_0)$  is directly equivalent to  $C_q$  and the imaginary part is directly equivalent to  $C_b$ . By virtue of the exponential term in (44), the contributions from skew quadrupole-like fields and axial fields can contribute to both the real and imaginary parts of  $C(\theta_0)$ . Thus contrary to what one might naturally expect from the simple theory, skew quadrupoles can be used to compensate solenoids, if the phasing of this exponential term is correctly managed. In a qualitative way, one can see that a plane inclined oscillation will turn into an elliptically polarised one, when the horizontal and vertical phase advances differ. This is taken into account by the  $(\mu_x - \mu_z)$  term in (44), which will typically oscillate +0.5 radian in a regular FODO lattice.

If one were to measure  $C_q$  ( $\text{Re}C$ ) and  $C_b$  ( $\text{Im}C$ ) as described earlier at a single point, it would not be possible to distinguish whether the machine was sinusoidally smooth or not. However, by measuring at several points this would become apparent, since although  $|C|$  would be constant the balance between  $\text{Re}C$  and  $\text{Im}C$  would oscillate.

Finally  $K(\theta)$  has a rather more general form than  $k(s)$  from (3) to take into account such situations as a solenoid with a slot in its end plates by which the beam enters and leaves. In this case, either  $\Delta B_x/\partial x$  or  $\partial B_z/\partial z$  could be zero. When calculating  $C(\theta_0)$  for a solenoid in a collider, the solenoid will most probably be inside a low- $\beta$  insertion. In this case, the

$$\frac{M(\theta)}{2} \left( \frac{\alpha_x}{\beta} - \frac{\alpha_z}{\beta} \right)$$

term can be important, and since  $\mu_x$  and  $\mu_z$  are varying rapidly in the low- $\beta$  region the choice of the solenoid length and the type of end plate (open or slot) can lead to situations with self-compensation.

## REFERENCES

- [1] G. Leleux, Proc. of Int. Symposium on e+e- Storage Rings, Saclay, Sept. 1966 (Presse Universitaire de France) pp. VII b-3-1/7.
- [2] P.L. Morton, Proc. of Int. Symposium on e+e- Storage Rings, Saclay, Sept. 1966 (Presse Universitaire de France) pp. VII b-1-7/5.
- [3] G. Ripken, Untersuchungen zur Strahlführung und Stabilität der Teilchenbewegung in Beschleunigern und Storage-Ringen unter strenger Berücksichtigung einer Kopplung der Betatronschwingungen , Int. Rep., DESY R1-70/4, June 1970.

- [4] G. Guignard, The general theory of all sum and difference resonances in a three-dimensional magnetic field in a synchrotron, CERN 76-06 (1976).
- [5] P.L. Morton, private communication.
- [6] P.J. Bryant, A simple theory for betatron coupling, CERN ISR-MA/75-28 (1975).
- [7] K. Takikawa, A simple and precise method for measuring the coupling coefficient of the difference resonance, CERN ISR-MA/75-34 (1975). See also, P. Bryant, P. Galbraith, J.-P. Gourber, G. Guignard and K. Takikawa, Measurement of the excitation of the coupling resonance  $Q_h - Q_v = \circ$ , CERN ISR-BOM/77-4 (1977) also published in IEEE Trans. N.S., Vol NS-24, no. 3, June 1977, pp. 1340-42.
- [8] P.J. Bryant and G. Guignard, Methods for measuring the complex coupling coefficient for the second order difference resonance  $Q_h = Q_v$  CERN ISR-MA/75-42 (1975).
- [9] J.-P. Koutchouk, Linear betatron coupling measurement and compensation in the ISR, CERN ISR-OP/80-27 (1980) also, published in Proc. XIth Int. Conf. on High Energy Accelerators, CERN, Geneva (Birkhauser Verlag, Basel, 1980) pp. 491-495.

# BEAM TRANSFER LINES

P.J. Bryant  
CERN, Geneva, Switzerland

## Abstract

After making some distinctions between transfer lines and circular machines, certain problems, typical of the type met by a transfer line designer, are discussed. The topics chosen include: steering, measurement of emittance and mismatch, setting tolerances for magnet alignment and excitation, emittance dilution due to mismatches and scattering in thin windows and lastly emittance-exchange insertions.

## 1. DISTINCTIONS BETWEEN TRANSFER LINES AND PERIODIC CIRCULAR MACHINES

Transmission of the position-velocity vector of a particle through a section of a transfer line, or circular machine, can be simply represented by a  $2 \times 2$  matrix (Fig. 1).

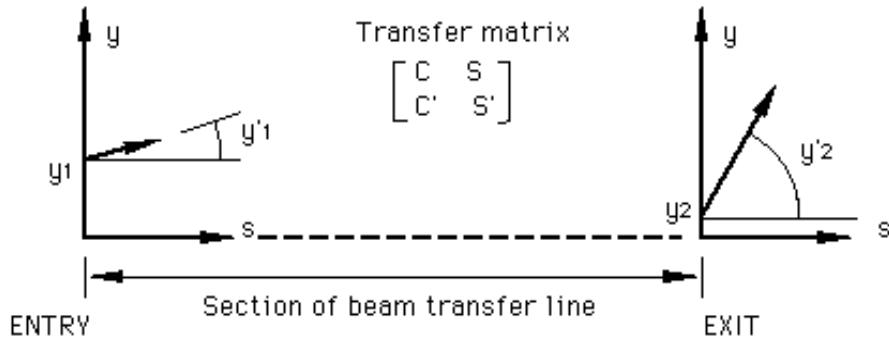


Fig. 1 Transmission through a section of lattice  
(y represents either transverse coordinate)

$$\begin{pmatrix} y_2 \\ y'_2 \end{pmatrix} = \begin{pmatrix} C & S \\ C' & S' \end{pmatrix} \begin{pmatrix} y_1 \\ y'_1 \end{pmatrix} = M_{1 \rightarrow 2} \begin{pmatrix} y_1 \\ y'_1 \end{pmatrix}. \quad (1)$$

The transfer matrix  $M_{1 \rightarrow 2}$  can be found by multiplying together the transfer matrices for the individual elements in the appropriate order. The individual matrices have the form,

$$M_y = \begin{pmatrix} \cos(\phi) & \frac{s}{\phi} \sin(\phi) \\ -\frac{\phi}{s} \sin(-\phi) & \cos(\phi) \end{pmatrix} \quad (2)$$

where

$$\phi = s\sqrt{|K|} \quad \text{and} \quad |K| = \left| \frac{1}{B\rho} \frac{dB_z}{dx} + \frac{1}{\rho^2} \right|$$

in accordance with the earlier lectures by J. Rossbach and P. Schmüser in these proceedings.

However, we often use a parameterized form for the matrix for a section of line, which was also given by J. Rossbach and P. Schmüser.

$$M_{1 \rightarrow 2} = \begin{pmatrix} \sqrt{\frac{\beta_2}{\beta_1}} [\cos \Delta\phi + \alpha_1 \sin \Delta\phi] & \sqrt{\beta_1 \beta_2} \sin \Delta\phi \\ -\frac{(1 + \alpha_1 \alpha_2) \sin \Delta\phi + (\alpha_2 - \alpha_1) \cos \Delta\phi}{\sqrt{\beta_1 \beta_2}} & \sqrt{\frac{\beta_1}{\beta_2}} [\cos \Delta\phi - \alpha_2 \sin \Delta\phi] \end{pmatrix} \quad (3)$$

In the first case of Eq. (1), the matrix is unambiguously determined, but in the second case of Eq. (3), there are in fact an infinite number of sets of parameters ( $\beta_1$ ,  $\beta_2$ ,  $\alpha_1$ ,  $\alpha_2$  and  $\Delta\phi$ ), which satisfy the numerical values of the matrix elements. This is the root of an important difference between circular machines and transfer lines, which sometimes leads to confusions.

## 1.1 Circular machines

A circular structure has an imposed periodicity, which imposes the same periodicity on the parameters  $\alpha$  and  $\beta$  and in fact *determines them uniquely*. If one samples the co-ordinates of an ion after each successive turn in a circular machine, the points will fill out an ellipse in phase space ( $y, y'$ ). Only one set of  $\alpha$  and  $\beta$  values fit that ellipse. It is the periodicity of the structure which makes it possible for that specific ellipse to be returned unchanged turn after turn and for this reason it is called the *matched ellipse* [Fig. 2(a)]. Now suppose one injects a beam of particles, whose spatial distribution defines a different ellipse characterized by some other parameters, say  $\alpha^*$  and  $\beta^*$ . The circular machine will not faithfully return this ellipse after each turn. Instead, the ellipse will tumble over and over filling out a much larger ellipse of the matched ellipse form [Fig. 2(b)]

In a truly linear system, the original ellipse will tumble round indefinitely inside the matched ellipse conserving its elliptical form and area, but in a practical system small non-linearities will cause an amplitude-frequency dependence, which will distort the ellipse. This is also shown in Fig. 2. Liouville's theorem requires the phase-space density to be conserved and in a strict mathematical sense this is true, since as the figure becomes more wound-up the spiral arms become narrower and the area is indeed constant. However it does not take long before the beam is apparently uniformly distributed over the matched ellipse and for all practical purposes the beam emittance has been increased. This is called *dilution of phase space by filamentation*, which is present to a greater or lesser extent at the injection into all circular machines.

Since filamentation will quickly transpose any beam ellipse into the matched ellipse in a circular machine, there is no point in using any  $\alpha$  and  $\beta$  values other than the matched ones.

Since  $\alpha$  and  $\beta$  depend on the whole structure any change at any point in the structure will in general (matched insertions excepted) change all the  $\alpha$  and  $\beta$  values everywhere.

## 1.2 Transfer lines

In a transfer line, there is no such restriction. The beam passes once and the shape of the ellipse at the entry to the line determines its shape at the exit. Exactly the *same* transfer line injected first with one emittance ellipse and then a different ellipse has to be accredited with *different*  $\alpha$  and  $\beta$  functions to describe the two cases. Thus  $\alpha$  and  $\beta$  depend on the input beam and their propagation depends on the structure. Any change in the structure will only change the  $\alpha$  and  $\beta$  values downstream of that point. There is an infinite number of sets of  $\alpha$  and  $\beta$  values, which can be used to describe the motion of a single ion in a transfer line (see Fig. 3)

and the choice of a particular set depends on the input ellipse shape. The input ellipse must be chosen by the designer and should describe the configuration of all the particles in the beam.

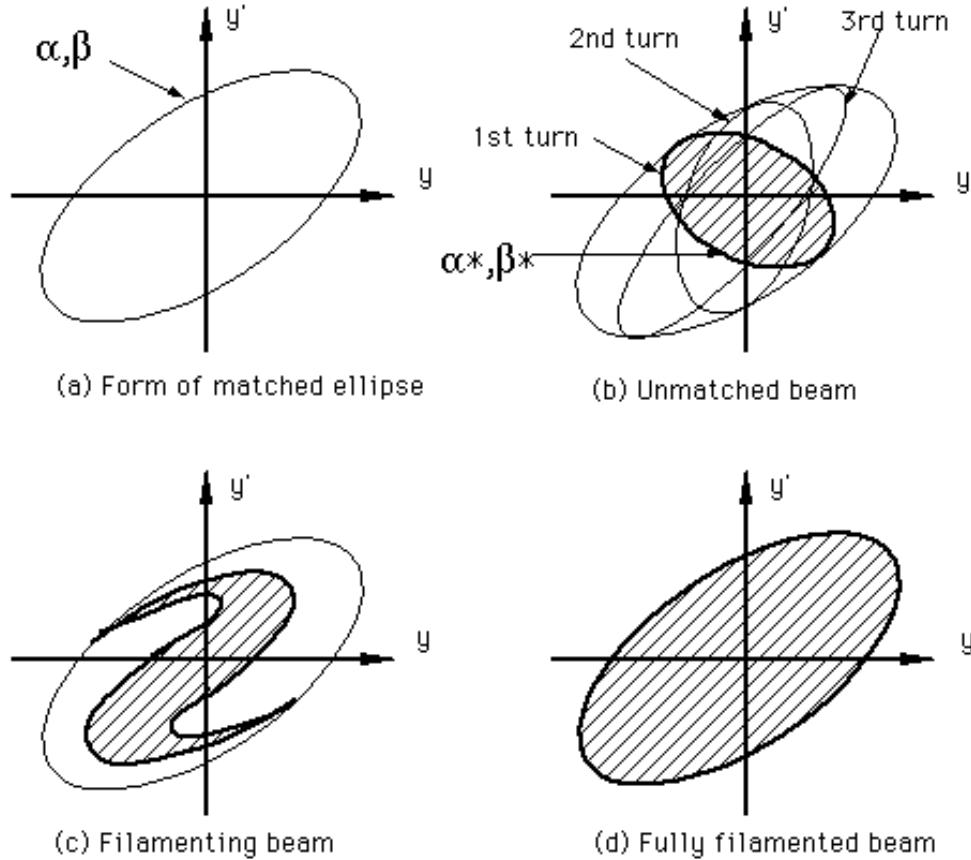


Fig. 2 Matched ellipse, unmatched and filamenting beam ellipses

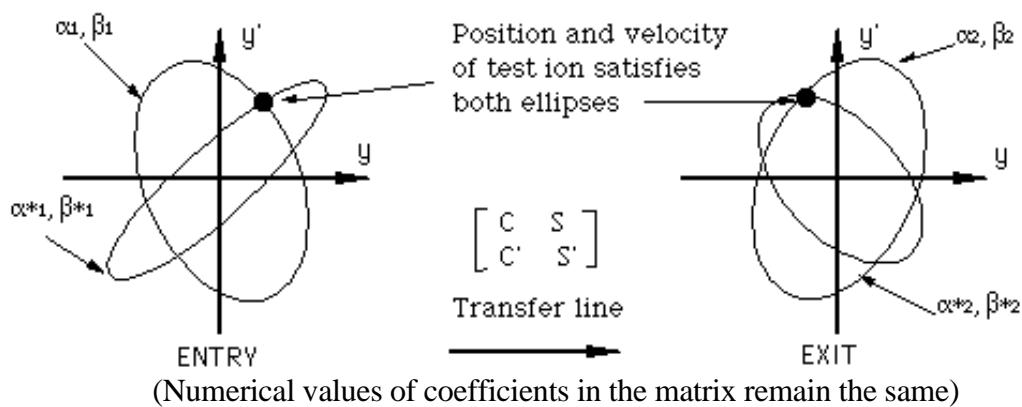


Fig. 3 Two ellipses from the infinite set that include the test ion

## 2. ORBIT CORRECTION IN TRANSFER LINES

Orbit correction, or steering, is basically straightforward in transfer lines, whereas in circular machines we could fill an entire course on the subject. The usual philosophy is illustrated in Fig. 4.

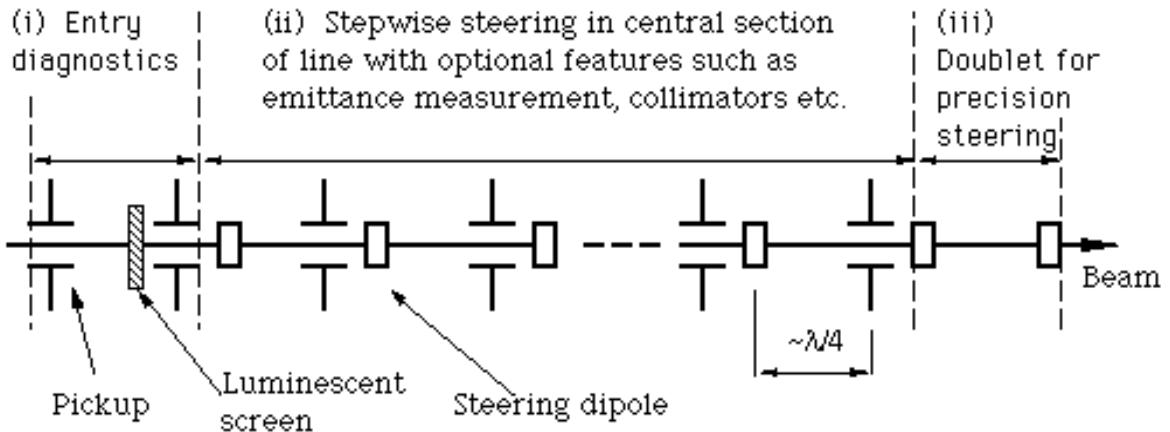


Fig. 4 Basic layout of diagnostic and correction elements for transfer line steering

- (i) At the entry to the line, it is useful to have a very clear diagnosis of beam position and angle and qualitative information on the shape, since this is usually the ejection from an accelerator and often a boundary of responsibility between groups. A pair of pickups and knowledge of the transfer matrix between them is in principle all that is needed to find the entry angle and position, but in practice, the precision and reliability of this measurement and its credibility as a diagnostic tool are greatly improved by having only a drift space between the pickups. The qualitative knowledge of the beam shape is most easily obtained with a luminescent screen and is of obvious diagnostic use.
- (ii) In the central section of the line, each steering magnet is paired with a pickup approximately a quarter of a betatron wavelength downstream, so that the trajectory can be corrected stepwise along the line. The direct application of this philosophy would lead to four pickups per betatron wavelength, but in practice, it is usual to find fewer pickups than this, especially if there are long straight sections. The measurement of beam emittance is usually made in the central part of the line in a dispersion-free section. The theory for the measurement of emittance and mismatches is treated in Section 4.
- (iii) At the exit to the line, the last two dipole correctors are used as a doublet to steer the beam to the angle and position, dictated by the closed orbit of the following accelerator or by a target. For maximum sensitivity, the dipoles should be approximately a quarter betatron wavelength apart.

The horizontal and vertical planes should be independent for correction elements. For example, tilted dipoles are sometimes used in the lattice of a transfer line, but correction coils for steering should be avoided on such magnets. Skew quadrupoles are occasionally used to interchange emittances between the horizontal and vertical planes. Such insertions also exchange the planes for steering. While being novel, this is quite acceptable, as long as no corrector is placed inside the skew quadrupole insertion, which would cause a coupling of its effect to both planes rather than a simple exchange.

Some care is needed in the positioning elements for the best sensitivity. The monitor controlling a steering magnet should be on the adjacent peak of the downstream beam oscillation (see Fig. 5), i.e. for the section of line from the steering dipole to the pickup, the matrix element  $S$  in Eq. (1) must be relatively large or in other terms  $\Delta\phi = \pi/2$  in Eq. (3).

The monitors and magnets should be sited near maxima in the  $\beta$ -function, since these are the most sensitive points for controlling and observing. This depends on the choice of input beam ellipse.

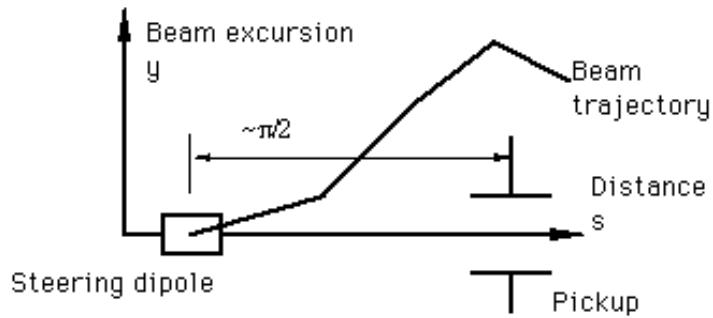


Fig. 5 Positioning of correction elements

Monitors can also be profitably placed in bends at points where off-momentum particles would have their maximum deviations. Using three well-placed pickups a bend can be used for momentum analysis. The simple linear matrices make the analysis of such systems very easy.

In a long line, a global correction may well be possible, followed by an exact beam steering at the end using two dipoles.

It should be possible to set the magnets in a transfer line and to be sure that the beam will be transmitted with 100% efficiency on the first try and that only a fine steering will be needed at the output to the line. If this is not the case, check that:

- the line is always cycled in the same way when it is powered and that the cycle saturates the magnets to set the hysteresis conditions.
- the current does not overshoot the requested value, especially when approaching the minimum value in the cycle and the final value. This is achieved by reducing the ramp rate when approaching set values.
- when a steering correction is made, that it is made using the standard excitation cycle. In this way, the value stored in the current file will reproduce the field exactly the next time the line is powered.
- check the position, angle and cross-section of the incoming beam.

Figure 6 shows computer output of the beam trajectory in the TT6-TT1 antiproton transfer line that was built in CERN. The first two pickups measure the incoming angle and position in both planes. These pickups are separated by exactly 10.75 m of free space. The next two pickups are in a long bend and act as a momentum analyser, in conjunction with an angle and position measurement made using the first two pickups. The remaining pickups have associated steering magnets. At the end of the line two dipoles match the beam to the ISR's closed orbit.

For the example shown in Fig. 6, it was found that a single corrector could virtually correct the whole trajectory with the result shown in Fig. 6. This type of correction is only practical with non-destructive pickups, which reliably record the complete trajectory in one shot, and an online computer for logging, display, analysis and application. The correction was stable and was applied throughout the life of the transfer line. The cycling of the magnets ensured that the beam reached the ISR on the first shot and only a fine-tuning of the injection was required for each new run. The TT6 line achieved 0.1mm accuracy with as little as  $10^9$  particles. All readings were logged and stored for later analysis and the detection of trends. The steering magnets were also equipped with Hall probes (temperature stabilized to +0.1 °C for outside ambient temperatures 15 °C to 34 °C). These probes made relative field changes extremely accurate, eliminating any hysteresis errors. This rather careful approach was justified by the scarcity of antiprotons and since setting-up could *not* be done with the reverse injection of protons.

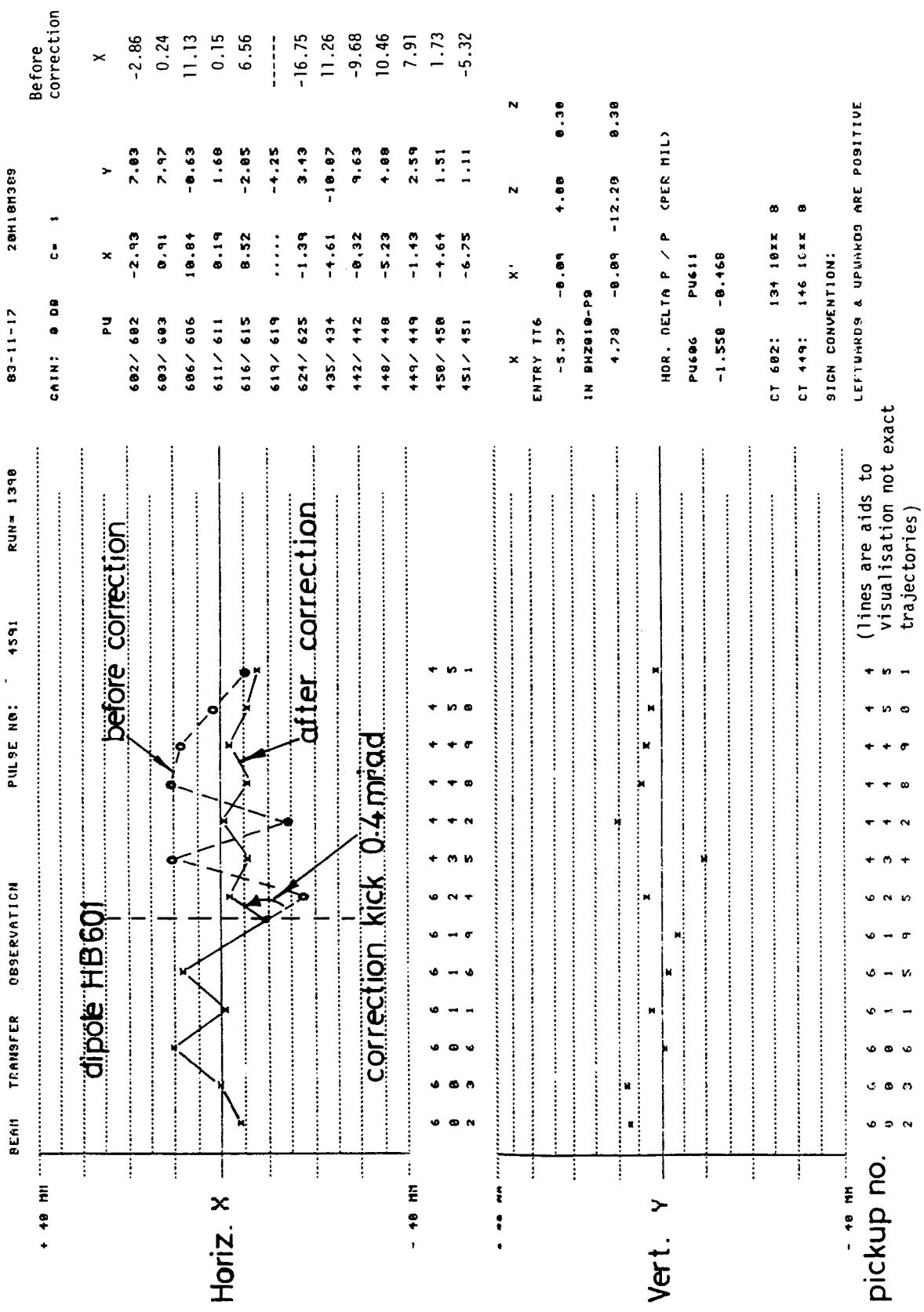


Fig. 6 Trajectories in the TT6-TT1 antiproton transfer line at CERN

### 3. MATCHING TRANSFER LINES

Ideally long transfer lines consist of a regular cell structure over the majority of their length with matching sections at either end to coordinate them with their injector and user machines. The regular part of the structure is then regarded as periodic and the simple FODO cell theory, given in earlier lectures by J. Rossbach and P. Schmüser, applies. Usually thin-lens formulae are quite sufficient. The matching sections are complicated and a complete course could be given on this. Basically one needs to match  $\beta$ ,  $\alpha$ ,  $D$ , and  $D'$  in both planes. In theory eight variables, that is eight quadrupole strengths and sometimes positions, need to be adapted. Some analytic solutions exist, but usually one uses a mixture of theory, intuition and computer optimization programmes.

### 4. EMITTANCE AND MISMATCH MEASUREMENT IN A DISPERSION-FREE REGION

With semi-destructive monitors, such as secondary emission grids or digitized luminescent screens, a density profile can be obtained of a beam. This profile is a projection of the population of the phase-space ellipse of the beam onto a transverse co-ordinate axis. In general, the profile is a near-Gaussian, but this is not really important for the following. From the profile, the standard deviation of the distribution,  $\sigma$  can be found and this can be used to define a beam width,  $W$ .  $W$  is then used to define the emittance  $\epsilon$ , but unfortunately several definitions are current.

$$\epsilon = \frac{W^2}{\beta} \pi = \begin{cases} \frac{(2\sigma)^2}{\beta} \pi & \text{Mostly used in proton machines, with or without } \pi \\ \frac{\sigma^2}{\beta} \pi & \text{Mostly used in electron machines, usually without } \pi \end{cases} \quad (4)$$

Somewhat arbitrarily,  $\epsilon = \sigma^2/\beta$  will be used in this paper.

If  $\beta$  is known unambiguously as in a circular machine, then a single profile measurement determines  $\epsilon$  by Eq. (4), but as can be understood from Section 1.2, it is not easy to be sure in a transfer line which  $\beta$  to use, or rather, whether the beam that has been measured is matched to the  $\beta$ -values used for the line. Indeed, the measurement of any mismatch is as important as the emittance itself. This problem can be resolved by using three monitors (see Fig. 7), i.e. the three width measurements determine the three unknowns  $\alpha$ ,  $\beta$  and  $\epsilon$  of the incoming beam.

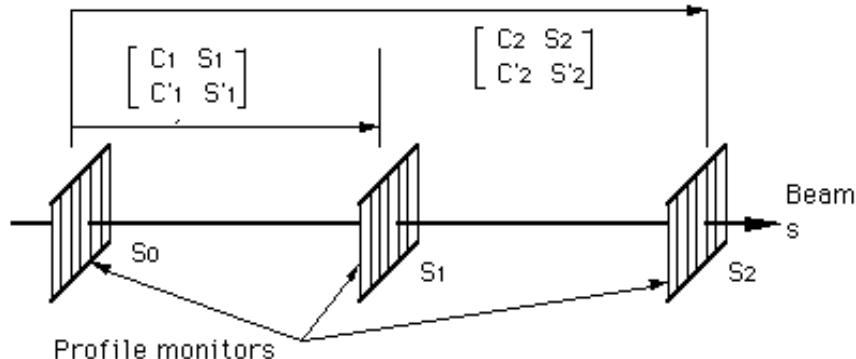


Fig. 7 Layout for emittance measurement

By definition, Eq. (4),

$$\varepsilon = \pi \frac{\sigma_0^2}{\beta_0} = \pi \frac{\sigma_1^2}{\beta_1} = \pi \frac{\sigma_2^2}{\beta_2} \quad (5)$$

where  $\beta_0$ ,  $\beta_1$  and  $\beta_2$  are the  $\beta$ -values corresponding to the beam and are therefore uncertain. Although we may not know  $\beta$  and  $\alpha$ , we do know the transfer matrices and how  $\beta$  and  $\alpha$  propagate through the structure (see lectures by K. Steffen in these proceedings).

$$\begin{pmatrix} \beta \\ \alpha \\ \gamma \end{pmatrix}_1 = \begin{pmatrix} C^2 & -2CS & S^2 \\ -CC' & CS' + SC' & -SS' \\ C'^2 & -2C'S' & S'^2 \end{pmatrix} \begin{pmatrix} \beta \\ \alpha \\ \gamma \end{pmatrix}_0 \quad (6)$$

where  $\gamma = (1 + \alpha^2)/\beta$ . Thus, from Eq. (6)

$$\beta_1 = C_1^2 \beta_0 - 2C_1 S_1 \alpha_0 + \frac{S_1^2}{\beta_0} (1 + \alpha_0^2) \quad (7)$$

$$\beta_2 = C_2^2 \beta_0 - 2C_2 S_2 \alpha_0 + \frac{S_2^2}{\beta_0} (1 + \alpha_0^2) \quad (8)$$

and from Eq. (5),

$$\beta_0 = \pi \frac{\sigma_0^2}{\varepsilon} \quad (9)$$

$$\beta_1 = \left( \frac{\sigma_1}{\sigma_0} \right)^2 \beta_0 \quad (10)$$

$$\beta_2 = \left( \frac{\sigma_2}{\sigma_0} \right)^2 \beta_0 \quad (11)$$

From Eqs. (7) and (8), we can find  $\alpha_0$  and using Eqs. (10) and (11), we can express  $\alpha_0$  as,

$$\alpha_0 = \frac{1}{2} \beta_0 \Gamma \quad (12)$$

where

$$\Gamma = \frac{(\sigma_2 / \sigma_0)^2 / S_2^2 - (\sigma_1 / \sigma_0)^2 / S_1^2 - (C_2 / S_2)^2 + (C_1 / S_1)^2}{(C_1 / S_1) - (C_2 / S_2)} . \quad (13)$$

Since  $\Gamma$  is fully determined, direct substitution back into Eq. (7) or Eq. (8), using Eq. (10) or Eq. (11) to re-express  $\beta_1$  or  $\beta_2$ , yields  $\beta_0$  which via Eq. (9) gives the emittance,

$$\beta_0 = 1 / \sqrt{\left( \frac{\sigma_2}{\sigma_0} \right)^2 / S_2^2 - \left( \frac{C_2}{S_2} \right)^2 + \left( \frac{C_1}{S_1} \right) \Gamma - \Gamma^2 / 4} \quad (14A)$$

$$\varepsilon = \left( \pi \sigma_0^2 \right) \sqrt{\left( \frac{\sigma_2}{\sigma_0} \right)^2 / S_2^2 - \left( \frac{C_2}{S_2} \right)^2 + \left( \frac{C_1}{S_1} \right) \Gamma - \Gamma^2 / 4} . \quad (14B)$$

The mismatch parameters  $\Delta\beta$  and  $\Delta\alpha$ , the differences between what is expected and what exists, can now be found directly from Eqs. (14B) and (12).

## 5. SMALL MISALIGNMENTS AND FIELD RIPPLE ERRORS IN DIPOLES AND QUADRUPOLES

One problem, which always faces a transfer line designer, is to fix the tolerances for magnet alignment and excitation currents. Although the following is rather idealistic and does not include such real-world problems as magnets having correlated ripple because they are on the same transformer, it does give a basis for fixing and comparing tolerances [1].

## 5.1 Dipole field and alignment errors in transfer lines

The motion of a particle in a transfer line can be written as

$$y = A\sqrt{\beta} \sin(\phi + B) \quad (15)$$

This motion is an ellipse in phase space with

$$y' = \frac{A}{\sqrt{\beta}} \cos(\phi + B) - \frac{A\alpha}{\sqrt{\beta}} \sin(\phi + B). \quad (16)$$

Rearranging we have

$$\begin{aligned} Y &= y / \sqrt{\beta} &= A \sin(\phi + B) \\ Y' &= y\alpha / \sqrt{\beta} + y'\sqrt{\beta} &= A \cos(\phi + B), \end{aligned} \quad (17)$$

where  $(Y, Y')$  are known as normalized phase-space coordinates since with these variables particles follow circular paths. Note that  $y'$  denotes  $dy/ds$ , while  $Y'$  denotes  $dY/d\phi$  and that  $\alpha = -1/2 d\beta/ds$ . The transformation to  $(Y, Y')$  is conveniently written in matrix form as

$$\begin{pmatrix} Y \\ Y' \end{pmatrix} = \begin{pmatrix} 1/\sqrt{\beta} & 0 \\ \alpha/\sqrt{\beta} & \sqrt{\beta} \end{pmatrix} \begin{pmatrix} y \\ y' \end{pmatrix}. \quad (18)$$

Consider now a beam for which the equi-density curves are circles in normalized phase space. If this beam receives an unwanted deflection,  $D$ , it will appear at the time of the deflection as shown in Fig. 8(a). However, this asymmetric beam distribution will not persist. As the beam continues along the transfer line, the particles will re-distribute themselves randomly in phase, while maintaining their distance from the origin, so as to restore rotational symmetry. This effect is known as filamentation (see also Section I.1). Thus after a sufficient time has elapsed the particles, which without the deflection  $D$  would have been at point P in Fig. 8(b), will be uniformly distributed at a radius  $D$  about the point P.

For one of these particles the projection onto the  $Y$ -axis will be

$$Y_2 = Y_1 + D \cos \theta,$$

where the subscripts 1 and 2 denote the unperturbed and perturbed positions respectively. Taking the square of this amplitude

$$Y_2^2 = Y_1^2 + 2Y_1 D \cos \theta + D^2 \cos^2 \theta$$

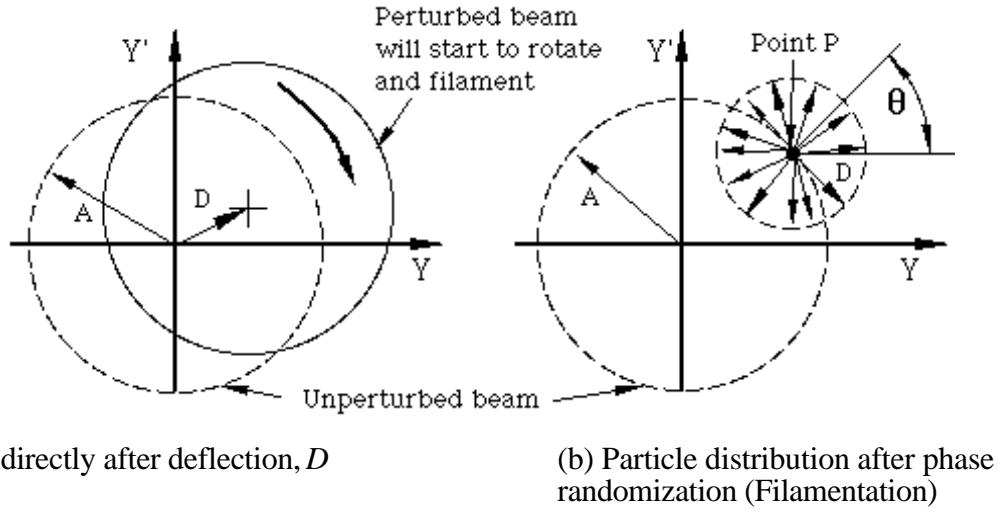


Fig. 8 Effect of an unwanted deflection

and then averaging over the particles around the point P after filamentation has randomized the kick gives\*

$$\langle Y_2^2 \rangle_p = \langle Y_1^2 \rangle_p + 2\langle Y_1 D \cos \theta \rangle_p + \langle D^2 \cos^2 \theta \rangle_p .$$

Since  $Y_1$  and  $D$  are uncorrelated (i.e.  $D$  does not depend on  $Y_1$ ), the second term can be written as

$$2\langle Y_1 D \cos \theta \rangle_p = 2\langle Y_1 \rangle_p \langle D \cos \theta \rangle_p .$$

The second factor is zero, since  $D$  is a constant [Fig. 8(a)], which gives,

$$\langle Y_2^2 \rangle_p = \langle Y_1^2 \rangle_p + \frac{1}{2} \langle D^2 \rangle_p = \langle Y_1^2 \rangle_p + \frac{1}{2} D^2 .$$

However, this result is true for any P at any radius A and hence it is true for the whole beam and

$$\langle Y_2^2 \rangle = \langle Y_1^2 \rangle + \frac{1}{2} D^2 . \quad (19)$$

Thus the emittance blow-up will be

$$\epsilon_2 = \epsilon_1 + \frac{\pi}{2} D^2 , \quad (20)$$

where, by definition (4),  $\epsilon = \pi \langle Y^2 \rangle$ , since  $Y = y/\sqrt{\beta}$  and  $\sigma^2 = \langle y^2 \rangle$ . The subscripts 1 and 2 refer to the unperturbed and perturbed emittances respectively. The expansion of the deflection,  $D$ , gives

$$D^2 = (\Delta Y)^2 + (\Delta Y')^2 = (\Delta y)^2 \frac{(1+\alpha^2)}{\beta} + (\Delta y')^2 \beta \quad (21)$$

so that (20) becomes

$$\epsilon_2 = \epsilon_1 + \frac{\pi}{2} \left[ (\Delta y)^2 \frac{(1+\alpha^2)}{\beta} + (\Delta y')^2 \beta \right] , \quad (22)$$

---

\*  $\langle \dots \rangle$  brackets indicate averaging over a distribution.

where  $\Delta y$  is a magnet alignment error and  $\Delta y' = l\Delta B/B\rho$  an angle error from a field error  $\Delta B$  of length  $l$ .

## 5.2 Gradient errors in transfer lines

Consider once again a beam for which the equi-density curves are circles in normalized phase space. If this beam sees a gradient error,  $k$ , the equi-density curves directly after the perturbation will be ellipses as shown in Fig. 9(a). Since the object of this analysis is to evaluate the effects of small errors, it is sufficient to regard this gradient error as a thin lens with the transfer matrix

$$\begin{pmatrix} y_2 \\ y'_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ k & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y'_1 \end{pmatrix} \quad (23)$$

where  $k = -l\Delta G/B\rho$  an amplitude-dependent kick due to a gradient error  $\Delta G$  of length  $l$ .

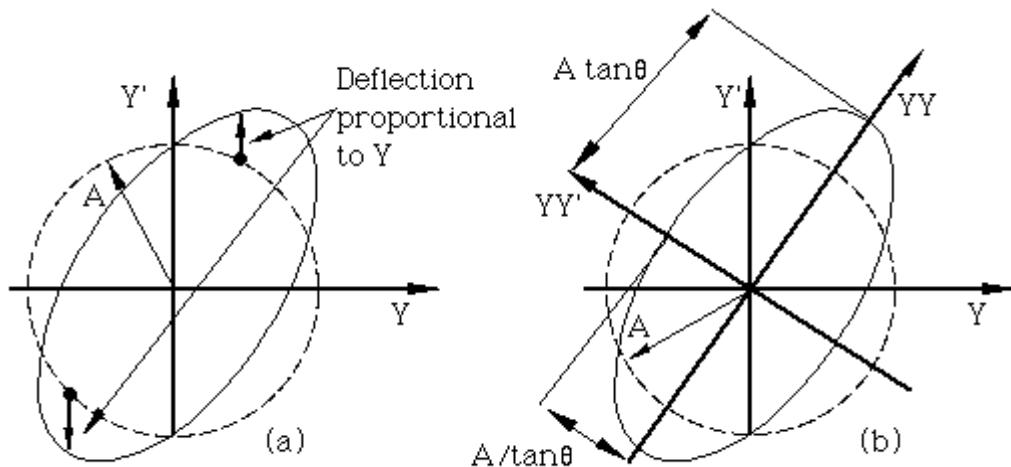


Fig. 9 Effect of a gradient error

Denoting the matrix in Eq. (18) as  $\mathbf{T}$ , it is easy to show that

$$\begin{pmatrix} Y_2 \\ Y'_2 \end{pmatrix} = \mathbf{T} \begin{pmatrix} 1 & 0 \\ k & 1 \end{pmatrix} \mathbf{T}^{-1} \begin{pmatrix} Y_1 \\ Y'_1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ k\beta & 1 \end{pmatrix} \begin{pmatrix} Y_1 \\ Y'_1 \end{pmatrix}.$$

It is now convenient to find a new co-ordinate system  $(YY, YY')$ , which is at an angle  $\theta$  to the  $(Y, Y')$  system, and in which the perturbed ellipse is a right ellipse [see Fig. 9(b)].

$$\begin{pmatrix} YY_2 \\ YY'_2 \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} 1 & 0 \\ k\beta & 1 \end{pmatrix} \begin{pmatrix} Y_1 \\ Y'_1 \end{pmatrix}. \quad (24)$$

If the initial distribution  $Y_1 = A \sin(\phi + B)$ ,  $Y'_1 = A \cos(\phi + B)$ , is introduced into the above expression, the new distribution will be

$$\begin{aligned} YY_2 &= A \sqrt{1 + k^2\beta^2 \sin^2\theta - 2k\beta \sin\theta \cos\theta} \sin(\phi + B + \Psi) \\ YY'_2 &= A \sqrt{1 + k^2\beta^2 \cos^2\theta + 2k\beta \sin\theta \cos\theta} \sin(\phi + B + \Psi'), \end{aligned} \quad (25)$$

where

$$\Psi = \tan^{-1} \left( \frac{-\sin \theta}{\cos \theta - k\beta \sin \theta} \right) \text{ and } \Psi' = \tan^{-1} \left( \frac{\cos \theta}{\sin \theta + k\beta \cos \theta} \right).$$

The  $(YY_2, YY'_2)$  ellipse will be a right ellipse when  $(\Psi - \Psi') = \pi/2$ , which gives the condition

$$\tan(2\theta) = 2/k\beta. \quad (26)$$

Equations (25) can be simplified using (26) and the relationship  $(\Psi - \Psi') = \pi/2$ . Equation (24) can then be rewritten as

$$\begin{pmatrix} YY_2 \\ YY'_2 \end{pmatrix} = \begin{pmatrix} \tan \theta & 0 \\ 0 & 1/\tan \theta \end{pmatrix} \begin{pmatrix} YY_1 \\ YY'_1 \end{pmatrix} \quad (27)$$

where

$$\begin{aligned} YY_1 &= A \sin(\phi + B') \\ YY'_1 &= A \cos(\phi + B') \\ B' &= B + \Psi = B + \tan^{-1}(1/\tan \theta). \end{aligned}$$

i.e.  $Y_1$  and  $Y'_1$  with a phase shift

Thus it has been possible to diagonalize Eq. (24) by introducing a phase shift  $\Psi$  into the initial distribution. Equation (27) is therefore not a true point-to-point transformation, as is Eq. (24) but since the initial distribution is rotationally symmetric the introduction of this phase shift has no effect.

The distance from the origin of a perturbed particle is given by Eq. (27) as

$$YY_2^2 + YY'^2 = A^2 \sin^2(\phi + B') \tan^2 \theta + A^2 \cos^2(\phi + B') \frac{1}{\tan^2 \theta}$$

Averaging over  $2\pi$  in  $\phi$  gives

$$\langle YY_2^2 + YY'^2 \rangle = \frac{1}{2} \left( \tan^2 \theta + \frac{1}{\tan^2 \theta} \right) \langle A^2 \rangle,$$

but

$$\langle A^2 \rangle = \langle YY_1^2 + YY'^2 \rangle = \langle Y_1^2 + Y'^2 \rangle$$

and from (26)

$$\tan^2 \theta + \frac{1}{\tan^2 \theta} = k^2 \beta^2 + 2.$$

Thus,

$$\langle YY_2^2 + YY'^2 \rangle = \frac{1}{2} (k^2 \beta^2 + 2) \langle YY_1^2 + YY'^2 \rangle. \quad (28)$$

As in the previous case for dipole errors, the asymmetric beam distribution will not persist. The beam will regain its rotational symmetry by filamentation. Each particle, however, will maintain its distance from the origin constant. Once filamentation has occurred, the distribution will not distinguish between the  $YY$  and  $YY'$  axes and Eq. (28) can be rewritten as

$$\langle YY_2^2 \rangle = \frac{1}{2} (k^2 \beta^2 + 2) \langle YY_1^2 \rangle \quad (29)$$

and hence the emittance blow-up will be

$$\epsilon_2 = \frac{1}{2} (k^2 \beta^2 + 2) \epsilon_1. \quad (30)$$

### 5.3 Combining errors

If there is a circular machine at the end of the transfer line, filamentation will take place there and the above expressions will give the emittance blow-up due to a *single* error in the preceding transfer line. A series of errors can be treated by taking them in beam order and assuming complete phase randomization between each error, although this is unlikely to be true in practical cases. By themselves, transfer lines are usually too short for the effects of filamentation to show and certainly there is never complete randomization between elements in a line. In the real world adjacent magnets are often on the same transformer, which also gives correlated errors. Having pointed out these deficiencies, the above method still gives a basis upon which to compare errors and fix tolerances. The assumption that full randomization takes place between elements will give a pessimistic result for the usual case of many independent elements, which errs on the correct side for fixing tolerances. For small numbers of elements with correlated errors however, the analysis may underestimate the effect.

## 6. EMITTANCE BLOW-UP DUE TO THIN WINDOWS IN TRANSFER LINES

Transfer lines are often built with a thin metal window separating their relatively poor vacuum from that of the accelerator or storage ring that they serve. The beam must pass through this window with as little degradation as possible. Luminescent screens are also frequently put into beams with the same hope that they will have a negligibly small effect on the beam emittance. It is therefore interesting to know how to calculate the blow-up for such cases.

The root mean square projected angle  $\theta_s$  due to multiple Coulomb scattering in a window is given by [2,3]

$$\sqrt{\langle \theta_s^2 \rangle} = \frac{0.0141}{\beta_c p [\text{MeV}/c]} Z_{\text{inc}} \sqrt{\frac{L}{L_{\text{rad}}}} \left( 1 + \frac{1}{9} \log_{10} \frac{L}{L_{\text{rad}}} \right) [\text{radian}] , \quad (31)$$

where  $Z_{\text{inc}}$  is particle charge in units of electron charge,  $p$  is the particle momentum in MeV/c,  $\beta_c = v/c$ ,  $L$  is thickness of scatterer and  $L_{\text{rad}}$  is radiation length of material of the scatterer.

Consider a particle with a projected angular deviation of  $y'_1$  at the window due to the initial beam emittance. This particle receives a net projected kick in the window of  $\theta_s$  and emerges with an angle  $y'_2$  given by

$$y'_2 = y'_1 + \theta_s .$$

By squaring and averaging over the whole beam this becomes

$$\langle y'^2 \rangle = \langle y'^2_1 \rangle + \langle \theta_s^2 \rangle + 2\langle y'_1 \theta_s \rangle$$

but, since  $y'_1$  is not correlated to  $\theta_s$ ,  $2\langle y'_1 \theta_s \rangle = 2\langle y'_1 \rangle \langle \theta_s \rangle = 0$  and the above simplifies to

$$\langle y'^2 \rangle = \langle y'^2_1 \rangle + \langle \theta_s^2 \rangle . \quad (32)$$

This describes the situation immediately after the scattering (see Fig. 10) when the beam is no longer matched. The position of the particles is unchanged since the scatterer is assumed to be thin.

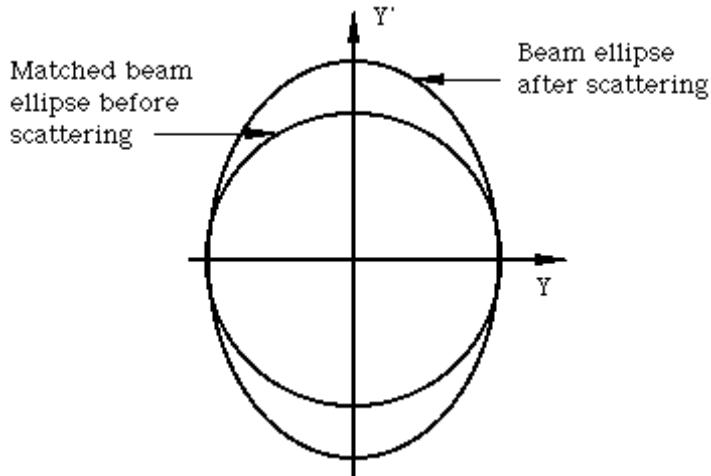


Fig. 10 Effect of a thin scatterer in normalized phase space

Using the same arguments as in Section 5.1, we see that this initial distribution filaments and the average angular divergence becomes

$$\langle y_2'^2 \rangle = \langle y_1'^2 \rangle + \frac{1}{2} \langle \theta_s^2 \rangle . \quad (33)$$

For conversion to emittance, the following relationship can be used,

$$\varepsilon_2 = \pi \left[ \frac{\alpha}{\beta} \langle y_2^2 \rangle + 2\alpha \langle y_2 y_2' \rangle + \beta \langle y_2'^2 \rangle \right] \quad (34)$$

which is found by re-writing (4) as  $\varepsilon = \pi \sigma^2 / \beta = \pi \langle Y^2 \rangle = \pi \langle Y'^2 \rangle$  and applying (17). The first term in (34) is unchanged by the scattering since the scatterer is assumed to be thin, so that  $\langle y_2^2 \rangle = \langle y_1^2 \rangle$ . The second term directly after the scattering yields,

$$\langle y_2 y_2' \rangle = \langle y_1 (y_1' + \theta_s) \rangle = \langle y_1 y_1' \rangle + \langle y_1 \theta_s \rangle$$

but since  $y_1$  and  $\theta_s$  are uncorrelated the second term can be written as  $\langle y_1 \rangle \langle \theta_s \rangle$  and is zero. Finally the third term can be evaluated by (33) after filamentation, so that,

$$\varepsilon_2 = \varepsilon_1 + \frac{\pi}{2} \beta \langle \theta_s^2 \rangle . \quad (35)$$

## 7. EMITTANCE DILUTION FROM BETATRON MISMATCH

This is basically the gradient error problem of Section 5.2 seen from a slightly different view point. It often happens that the constraints on the linear optics are such that an analytically perfect match cannot be found between the end of a transfer line and the accelerator it serves. It may also be that measurements of the beam ellipse reveal a mismatch of unknown origin. These situations pose the problem of what error in  $\beta$  and  $\alpha$  can be tolerated? The designer must therefore be able to convert the mismatch into an emittance increase that can be judged against criteria such as the acceptable loss of luminosity in a collider.

The transformation of the phase-space motion to normalised coordinates  $(Y, Y')$  was given in (18). If this transformation is prepared for an ellipse characterised by  $\alpha_1$  and  $\beta_1$ , but is then applied to a mismatched ellipse characterised by  $\alpha_2$  and  $\beta_2$  [using (15) and (16)] as indicated below,

$$\begin{pmatrix} Y \\ Y' \end{pmatrix} = \begin{pmatrix} 1/\sqrt{\beta_1} & 0 \\ \alpha_1/\sqrt{\beta_1} & \sqrt{\beta_1} \end{pmatrix} \begin{pmatrix} y \\ y' \end{pmatrix} = \begin{pmatrix} 1/\sqrt{\beta_1} & 0 \\ \alpha_1/\sqrt{\beta_1} & \sqrt{\beta_1} \end{pmatrix} \begin{pmatrix} A\sqrt{\beta_2} \sin(\phi + B) \\ \frac{A}{\sqrt{\beta_2}} \cos(\phi + B) - \frac{A\alpha_2}{\sqrt{\beta_2}} \sin(\phi + B) \end{pmatrix} \quad (36)$$

then an ellipse is obtained in the normalised phase space (see Fig. 11) with the equation<sup>†</sup>,

$$Y^2 \left[ \frac{\beta_1}{\beta_2} + \left( \alpha_1 - \alpha_2 \frac{\beta_1}{\beta_2} \right)^2 \frac{\beta_2}{\beta_1} \right] + Y'^2 \frac{\beta_2}{\beta_1} - 2YY' \left( \alpha_1 - \alpha_2 \frac{\beta_1}{\beta_2} \right) \frac{\beta_2}{\beta_1} = A^2 \quad (37)$$

that can be compared to the circle given by the matched beam

$$Y^2 + Y'^2 = A^2 \quad (38)$$

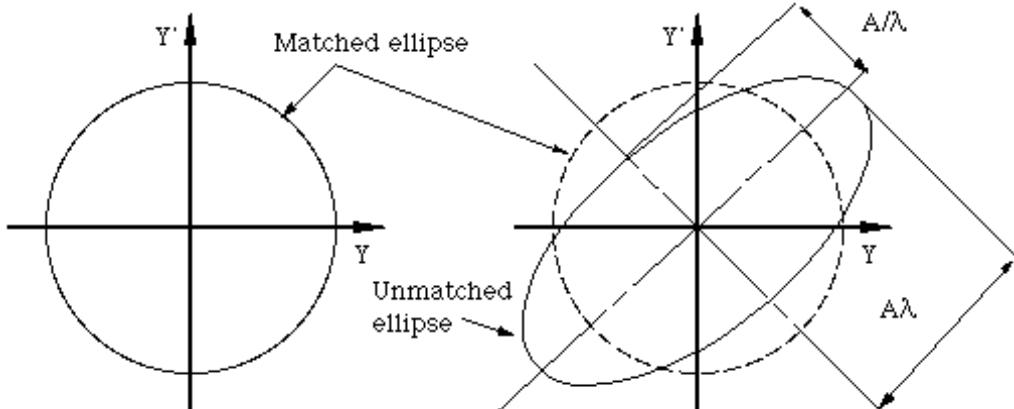


Fig. 11 Betatron mismatch in normalised phase space

Equation (37) is exactly similar in form to a general phase space ellipse in normal space and if we apply the equivalents,

$$\gamma \equiv \frac{\beta_1}{\beta_2} + \left( \alpha_1 - \alpha_2 \frac{\beta_1}{\beta_2} \right)^2 \frac{\beta_2}{\beta_1}, \quad \beta \equiv \frac{\beta_2}{\beta_1} \quad \text{and} \quad \alpha \equiv - \left( \alpha_1 - \alpha_2 \frac{\beta_1}{\beta_2} \right) \frac{\beta_2}{\beta_1} \quad (39)$$

all the standard formulae [4] can be used. One can easily check for example that  $\gamma = (1 + \alpha^2)/\beta$  still holds. Thus, we can avoid a lot of tedious algebra and quote directly the major and minor axes,  $a$  and  $b$  of the mismatched ellipse,

$$a = \frac{A}{\sqrt{2}} (\sqrt{H+1} + \sqrt{H-1}) \quad \text{and} \quad b = \frac{A}{\sqrt{2}} (\sqrt{H+1} - \sqrt{H-1}) \quad (40)$$

---

<sup>†</sup> Multiply out (36) and then eliminate the sine and cosine terms by squaring and adding the two equations for  $Y$  and  $Y'$ .

where

$$H = \frac{1}{2} \left[ \frac{\beta_1}{\beta_2} + \left( \alpha_1 - \alpha_2 \frac{\beta_1}{\beta_2} \right)^2 \frac{\beta_2}{\beta_1} + \frac{\beta_2}{\beta_1} \right]. \quad (41)$$

As in Section 5.2, the circle of the matched beam can be converted to the ellipse of the mismatched beam by the application of a diagonal matrix of the form

$$\begin{pmatrix} \lambda & 0 \\ 0 & 1/\lambda \end{pmatrix}$$

after a suitable rotation. The rotation has no significant influence since the original distribution is rotationally symmetric. From (40) we see that

$$\lambda = \frac{1}{\sqrt{2}} (\sqrt{H+1} + \sqrt{H-1}) \text{ and } \frac{1}{\lambda} = \frac{1}{\sqrt{2}} (\sqrt{H+1} - \sqrt{H-1}). \quad (42)$$

It is quickly verified that the two equations in (42) are consistent. Thus the square of the distance of a particle from the origin is

$$Y^2 + Y'^2 = \lambda^2 A^2 \sin^2(\phi + B) + \lambda^{-2} A^2 \cos^2(\phi + B). \quad (43)$$

Averaging over all phases simplifies (43) to

$$\langle Y^2 + Y'^2 \rangle = \frac{1}{2} (\lambda^2 + \lambda^{-2}) A^2. \quad (44)$$

Since the factor  $(\lambda^2 + \lambda^{-2})$  is independent of radius and orientation, (44) applies to the whole beam independent of its distribution. Thus, we can express the effective increase in emittance as,

$$\epsilon_{\text{diluted}} = \frac{1}{2} \left( \lambda^2 + \frac{1}{\lambda^2} \right) \epsilon_0. \quad (45)$$

First we note that even for quite large values of  $\lambda$ , the effect on the emittance is less than one might intuitively expect. For example, if  $\lambda=1.4$  the circle in normalised phase space is deformed in the ratio 2:1 and yet the emittance is only increased by 23%. Evaluating  $(\lambda^2 + \lambda^{-2})/2$  by use of (42), we find that

$$\epsilon_{\text{diluted}} = H \epsilon_0 = \frac{1}{2} \left[ \frac{\beta_1}{\beta_2} + \left( \alpha_1 - \alpha_2 \frac{\beta_1}{\beta_2} \right)^2 \frac{\beta_2}{\beta_1} + \frac{\beta_2}{\beta_1} \right] \epsilon_0 \quad (46)$$

where the subscript 1 denotes the expected values for the beam ellipse and subscript 2 denotes the mismatched values.

## 8 . EMITTANCE EXCHANGE INSERTION

Beams usually have different emittances in the two transverse planes and it can happen that there is a preference for having the smaller value in a particular plane. For example, in a

collider with a horizontal crossing angle the luminosity is independent of the horizontal emittance and it is therefore an advantage to arrange for the smaller of the two emittances to be in the vertical plane. The exchange of the emittances can be made in the transfer line before injection to the collider. A complete exchange of the transverse phase planes requires a transformation of the form,

$$\begin{pmatrix} x \\ x' \\ z \\ z' \end{pmatrix}_2 = \begin{pmatrix} 0 & 0 & m_{13} & m_{14} \\ 0 & 0 & m_{23} & m_{24} \\ m_{31} & m_{32} & 0 & 0 \\ m_{41} & m_{42} & 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ x' \\ z \\ z' \end{pmatrix}_1 . \quad (47)$$

This can be achieved by using skew quadrupole lenses. First we shall derive a thin-lens formulation for a skew quadrupole and then search for the conditions to satisfy the above transformation.

For a rotated co-ordinate system (see Fig. 12), the rotation matrix,  $\mathbf{R}$ , is given by

$$xx = x \cos \theta + y \sin \theta \quad \text{so that} \quad \mathbf{R} = \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix} . \quad (48)$$

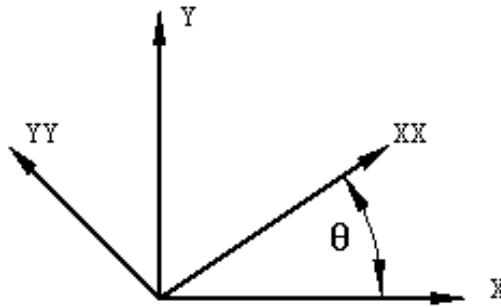


Fig. 12 Rotated co-ordinate system

The skew quadrupole lens is just a normal lens rotated by  $45^\circ$ . Thus the transfer matrix,  $\mathbf{M}_s$ , in the thin-lens approximation would be related to the transfer matrix,  $\mathbf{M}_q$ , of the normal lens by,

$$\mathbf{M}_s = \mathbf{R}^{-1} \mathbf{M}_{\text{quad}} \mathbf{R} \quad (49)$$

$$\mathbf{M}_s = \begin{pmatrix} \cos \theta & 0 & -\sin \theta & 0 \\ 0 & \cos \theta & 0 & -\sin \theta \\ \sin \theta & 0 & \cos \theta & 0 \\ 0 & \sin \theta & 0 & \cos \theta \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ \delta_s & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & -\delta_s & 1 \end{pmatrix} \begin{pmatrix} \cos \theta & 0 & \sin \theta & 0 \\ 0 & \cos \theta & 0 & \sin \theta \\ -\sin \theta & 0 & \cos \theta & 0 \\ 0 & -\sin \theta & 0 & \cos \theta \end{pmatrix}$$

$$\mathbf{M}_s = \begin{pmatrix} 1 & 0 & 0 & 0 \\ \delta_s \cos 2\theta & 1 & \delta_s \sin 2\theta & 0 \\ 0 & 0 & 1 & 0 \\ \delta_s \sin 2\theta & 0 & -\delta_s \cos 2\theta & 1 \end{pmatrix}$$

For a pure skew quadrupole  $\theta = \pi/4$ ,  $\sin 2\theta = 1$  and  $\cos 2\theta = 0$

$$\mathbf{M}_s = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & \delta_s & 0 \\ 0 & 0 & 1 & 0 \\ \delta_s & 0 & 0 & 1 \end{pmatrix}. \quad (50)$$

Consider now three skew quadrupoles  $\delta_1$ ,  $\delta_2$ , and  $\delta_3$  in a normal lattice (see Fig. 13).

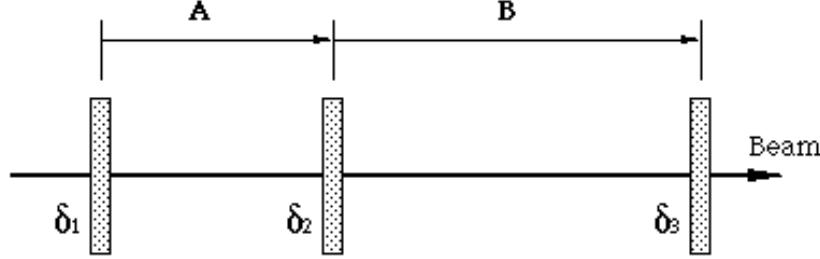


Fig. 13 Three skew quadrupoles embedded in a normal lattice

The normal lattice matrix is  $\mathbf{C} = \mathbf{BA}$ , and the lattice with skews is,  $\mathbf{M} = (\delta_3)\mathbf{B}(\delta_2)\mathbf{A}(\delta_1)$ . To save you the bother of matrix multiplication we quote the result,

$$\mathbf{M} = \left( \begin{array}{|c|c|c|c|} \hline & c_{12} & (c_{12}\delta_1 + b_{12}a_{33}\delta_2) & b_{12}a_{34}\delta_2 \\ \hline & \left[ \begin{array}{l} c_{21} + b_{22}a_{34}\delta_1\delta_2 \\ +\delta_3(c_{34}\delta_1 + b_{34}a_{11}\delta_2) \end{array} \right] & \left[ \begin{array}{l} c_{22}\delta_1 + b_{22}a_{33}\delta_2 \\ +\delta_3(c_{33} + b_{34}a_{12}\delta_1\delta_2) \end{array} \right] & \left( \begin{array}{l} b_{22}a_{34}\delta_2 \\ +\delta_3c_{34} \end{array} \right) \\ \hline & b_{34}a_{12}\delta_2 & (c_{33} + b_{34}a_{12}\delta_1\delta_2) & c_{34} \\ \hline & \left[ \begin{array}{l} \delta_3(c_{11} + b_{12}a_{34}\delta_1\delta_2) \\ +c_{44}\delta_1 + b_{44}a_{11}\delta_2 \end{array} \right] & \left[ \begin{array}{l} \delta_3(c_{12}\delta_1 + b_{12}a_{33}\delta_2) \\ +c_{43} + b_{44}a_{12}\delta_1\delta_2 \end{array} \right] & \left( \begin{array}{l} b_{12}a_{34}\delta_2\delta_3 \\ +c_{44} \end{array} \right) \\ \hline \end{array} \right) \quad (51)$$

In order to force (51) into the form of (47), we need the top left-hand side and bottom right-hand side sub-matrices to be zero. Thus,

$$\begin{aligned} 0 &= c_{11} + b_{12}a_{34}\delta_1\delta_2 & 0 &= c_{33} + b_{34}a_{12}\delta_1\delta_2 \\ 0 &= c_{21} + b_{22}a_{34}\delta_1\delta_2 + \delta_3(c_{34}\delta_1 + b_{34}a_{11}\delta_2) & 0 &= \delta_3(c_{12}\delta_1 + b_{12}a_{33}\delta_2) + c_{43} + b_{44}a_{12}\delta_1\delta_2 \\ 0 &= c_{12} & 0 &= c_{34} \\ 0 &= c_{22} + b_{34}a_{12}\delta_2\delta_3 & 0 &= b_{12}a_{34}\delta_2\delta_3 + c_{44} \end{aligned}$$

The most basic requirements are,  $c_{12} = c_{34} = 0$ . Thus, the basic lattice should give,

$$\Delta\phi_x = n\pi \quad \text{and} \quad \Delta\phi_z = m\pi \quad (n, m \text{ integer}). \quad (52)$$

For example, an [FDFD] structure with 90° phase advance per cell, or an [FDFDFD] with 60° phase advance per cell would be suitable. Assuming that this phase condition is satisfied, then we can write for the lens strengths  $\delta_1$ ,  $\delta_2$  and  $\delta_3$

$$\delta_1\delta_2 = -\frac{c_{11}}{a_{34}b_{12}} = -\frac{c_{33}}{a_{12}b_{34}} \quad (53)$$

$$\delta_2\delta_3 = -\frac{c_{22}}{a_{12}b_{34}} = -\frac{c_{44}}{a_{34}b_{12}} \quad (54)$$

Equations (53) and (54) indicate that symmetry is needed between the horizontal and vertical planes, which can be satisfied by the FODO cells mentioned above. Despite the apparent complexity of the coefficients it is possible to find solutions. For example let us choose an [FDFD] structure with 90 ° per cell and the skew quadrupoles set at symmetric positions as shown in Fig. 14.

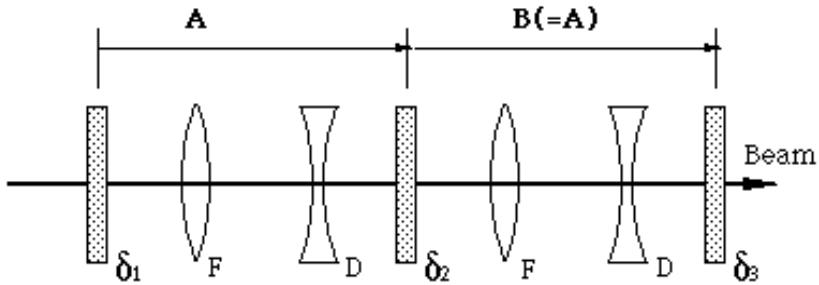


Fig. 14 Example of insertion for phase-plane exchange

By symmetry  $\mathbf{A} = \mathbf{B}$  and the input  $\alpha$  and  $\beta$  values will equal the output values and also the values at the central skew quadrupole. So,

$$\mathbf{A}_z \left(= \mathbf{B}_z\right) = \begin{pmatrix} \left(\cos \Delta\phi_z + \alpha_z \sin \Delta\phi_z\right) & \beta_z \sin \Delta\phi_z \\ -\left(1 + \alpha_z^2\right)\beta_z^{-1} \sin \Delta\phi_z & \left(\cos \Delta\phi_z - \alpha_z \sin \Delta\phi_z\right) \end{pmatrix}$$

Matrix  $\mathbf{C}$  will have a similar form with the phase shifts  $2\Delta\phi_x$  and  $2\Delta\phi_z$ . For this case we have chosen  $\Delta\phi_x = \Delta\phi_z = \pi/2$ , so that

$$\mathbf{A} = \mathbf{B} = \begin{pmatrix} \alpha_x & \beta_x & | 0 & 0 \\ -\frac{(1+\alpha_x^2)\beta_x^{-1}}{\alpha_x} & -\alpha_x & | 0 & 0 \\ \hline 0 & 0 & | \alpha_z & \beta_z \\ 0 & 0 & | -(1+\alpha_z^2)\beta_z^{-1} & -\alpha_z \end{pmatrix} \quad (55)$$

$$\mathbf{C} = \begin{pmatrix} -1 & 0 & | 0 & 0 \\ 0 & -1 & | 0 & 0 \\ \hline 0 & 0 & | -1 & 0 \\ 0 & 0 & | 0 & -1 \end{pmatrix} \quad (56)$$

The skew quadrupole gradients would then be,

$$\delta_1 = \delta_2 = \delta_3 = \frac{1}{\sqrt{\beta_x \beta_z}} \quad (57)$$

Referring back to the thin-lens formulation of a FODO cell in the lectures by J. Rossbach and P. Schmüser in these proceedings, we see that

$$\sin\left(\frac{\Delta\phi}{2}\right) = \frac{L}{2} \delta_F = -\frac{L}{2} \delta_D$$

and for  $\Delta\phi = \pi/2$

$$\delta_F = -\delta_D = \frac{\sqrt{2}}{L} \quad (58)$$

where  $L$  is the half-cell length. Other examples can be found in Ref. [5].

\* \* \*

## REFERENCES

- [1] H.G. Hereward, "How Good is the R.M.S. as a Measure of Beam Size?", CERN/MPS/DC 69-15 (1969).
- [2] Particle Physics Group, Review of Particle Properties, CERN, Geneva, August 1982 (reprinted from Physics Letters, Vol. 111B, April 1982).
- [3] E. Fischer, private communication, (ISR-VA/EF/sm, Beam blow-up due to multiple scattering in the injection windows, 25 Jan. 1974, Internal Note).
- [4] C. Bovet, R. Gouiran, I. Gumowski, K.H. Reich, A selection of formulae and data useful for the design of A.G. synchrotrons, CERN/MPS-SI/Int. DL/70/4, (April 1970).
- [5] L.R. Evans, A phase plane exchange section for the SPS antiproton injection beam line, CERN/SPS-DI(MST)/80-2.

# NON-LINEARITIES AND RESONANCES

*E. Wilson*  
CERN, Geneva, Switzerland

## 1. INTRODUCTION

The first step in designing an accelerator or storage ring is to choose an optimum pattern of focusing and bending magnets, the lattice. At this stage, non-linearities in the guide field are ignored. It is assumed that the bending magnets are identical and have a pure dipole field. Gradient magnets or quadrupoles have radial field shapes which have a constant slope, unperturbed by higher-order multipole terms.

Before going too far in fixing parameters, the practical difficulties in designing the magnets must be considered and the tolerances which can be reasonably written into the engineering specification determined. Estimates must be made of the non-linear departures from pure dipole or gradient field shape, and of the statistical fluctuation of these errors around the ring at each field level.

We must take into consideration that the remanent field of a magnet may have quite a different shape from that defined by the pole geometry; that steel properties may vary during the production of laminations; that eddy currents in vacuum chamber and coils may perturb the linear field shape. Mechanical tolerances must ensure that asymmetries do not creep in. At high field the linearity may deteriorate owing to saturation and variations in packing factor can become important. Superconducting magnets will have strong error fields due to persistent currents in their coils.

When these effects have been reviewed, tolerances and assembly errors may have to be revised and measures taken to mix or match batches of laminations with different steel properties or coils made from different batches of superconductor. It may be necessary to place magnets in a particular order in the ring in the light of production measurements of field uniformity or to shim some magnets at the edge of the statistical distribution. Even when all these precautions have been taken, non-linear errors may remain whose effect it is simpler to compensate with auxiliary multipole magnets.

Apart from the obvious need to minimize closed orbit distortion, these measures must be taken to reduce the influence of non-linear resonances on the beam. A glance at the working diagram (Fig. 1) shows why this is so. The  $Q_H$ ,  $Q_V$  plot is traversed by a mesh of non-linear resonance lines or stopbands of first, second, third, and fourth order. The order,  $n$ , determines the spacing in the  $Q$  diagram; third-order stopbands, for instance, converge on a point which occurs at every  $1/3$  integer  $Q$ -value (including the integer itself). The order,  $n$ , is related to the order of the multipole which drives the resonance. For example, fourth-order resonances are driven by multipoles with  $2n$  poles, i.e. octupoles. Multipoles can drive resonances of lower-order; octupoles drive fourth- and second-order; sextupoles third- and first-order, etc., but here we simply consider the highest order driven.

The non-linear resonances are those of third-order and above, driven by non-linear multipoles. Their strength is amplitude-dependent so that they become more important as we seek to use more and more of the machine aperture. Theory used to discount resonances of fifth- and higher-order as harmless (self-stabilized), but experience in the ISR, FNAL and SPS suggests this is not to be relied upon when we want beams to be stored for more than a second or so.

Each resonance line is driven by a particular pattern of multipole field error which can be present in the guide field. The lines have a finite width depending directly on the strength of the

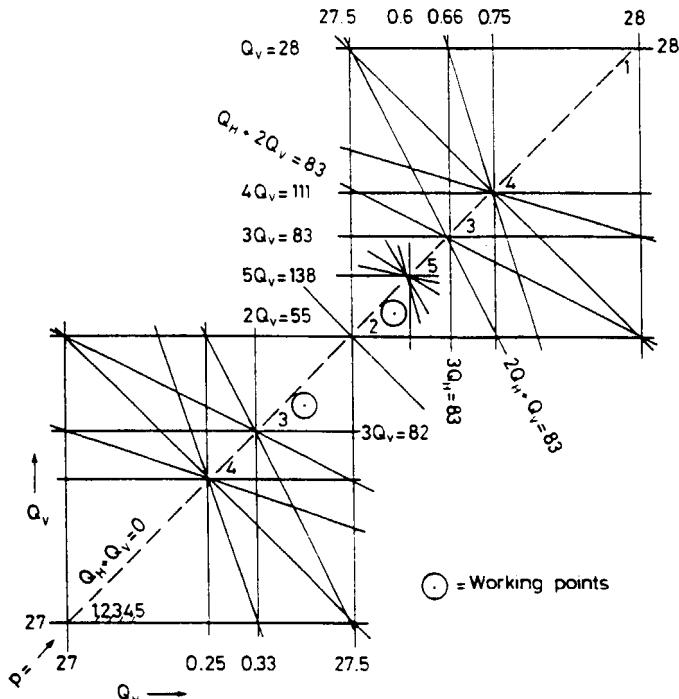


Fig. 1 Working diagram or  $Q_H$ ,  $Q_V$  plot showing the non-linear resonances in the operating region of the CERN SPS

error. In the case of those driven by non-linear fields, the width increases as we seek to exploit a larger fraction of the magnet aperture. We must ensure that the errors are small enough to leave some clear space between the stopbands to tune the machine, otherwise particles will fall within the stopbands and be rapidly ejected before they have even been accelerated. In general, the line width is influenced by the random fluctuations in multipole error around the ring rather than the mean multipole strength.

Systematic or average non-linear field errors also make life difficult. They cause  $Q$  to be different for the different particles in the beam depending on their betatron amplitude or momentum defect. Such a  $Q$ -spread implies that the beam will need a large resonance-free window in the  $Q$  diagram. In the case of the large machines, SPS, LEP, HERA, etc., the window would be larger than the half integer square itself if we did not balance out the average multipole component in the ring by powering correction magnets.

Paradoxically, when a "pure" machine has been designed and built, there is often a need to impose a controlled amount of non-linearity to correct the momentum dependence of  $Q$  or to introduce a  $Q$ -spread among the protons to prevent a high intensity instability. Sextupole and octupole magnets may have to be installed to this end and techniques studied which will enable the control room staff to find the correct settings for these trim magnets once the machine works.

Yet another set of multipole magnets is often required in a pulsed synchrotron to deliberately excite non-linear betatron resonances and extract the beam in a long slow spill.

With modern computer control, the correction of closed orbit distortion due to linear field errors has become a routine matter and, particularly in large accelerators, most of the emphasis has shifted to calculation and elimination of the non-linear effects which prove to be of considerable importance in the running-in of FNAL and the SPS. In this talk I hope to outline sufficient of the physics and mathematics of non-linearities to introduce the reader to this important aspect of accelerator theory.

## 2. MULTIPOLE FIELDS

Before we come to discuss the non-linear terms in the dynamics, we shall need to describe the field errors which drive them. The magnetic *vector potential* of a magnet with  $2n$  poles in Cartesian coordinates is:

$$A = \sum A_n f_n(x, z) , \quad (1)$$

where  $f_n$  is a homogeneous function in  $x$  and  $z$  of order  $n$ .

**Table 1**  
Cartesian solutions of magnetic vector potential

Multipole	$n$	Regular $f_n$	Skew $f_n$
Quadrupole	2	$x^2 - z^2$	$2xz$
Sextupole	3	$x^3 - 3xz^2$	$3x^2z - z^3$
Octupole	4	$x^4 - 6x^2z^2 + z^4$	$4x^3z - 4xz^3$
Decapole	5	$x^5 - 10x^3z^2 + 5xz^4$	$5x^4z - 10x^2z^3 + z^5$

Table 1 gives  $f_n(x, z)$  for low-order multipoles, both regular and skew. Figure 2 shows the distinction. We can obtain the function for other multipoles from the binomial expansion of

$$f_n(x, z) = (x + iz)^n . \quad (2)$$

The real terms correspond to regular multipoles, the imaginary ones to skew multipoles.

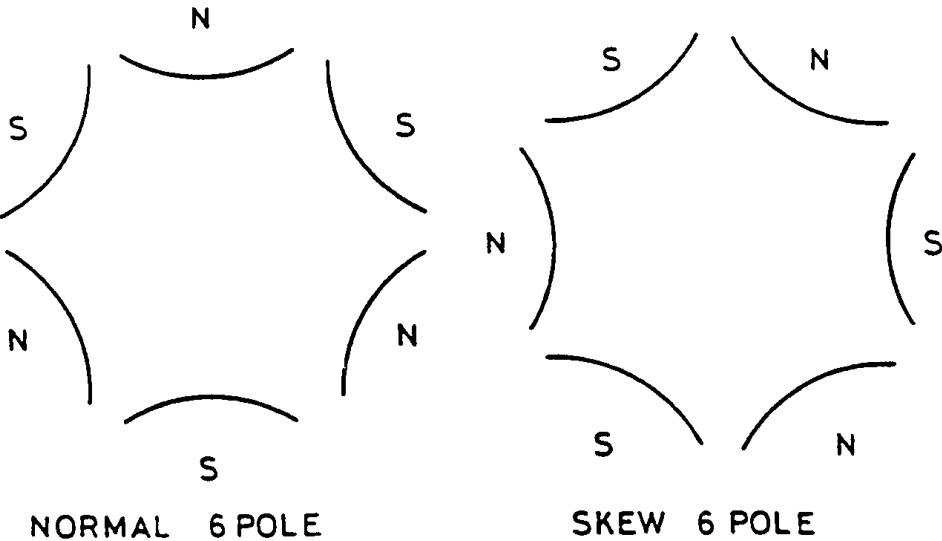


Fig. 2 Pole configurations for a regular sextupole and a skew sextupole

For numerical calculations it is useful again to relate  $A_n$  and field, remembering that for regular magnets:

$$B_z(z=0) = \frac{\partial A_s}{\partial x} = \sum_{n=1}^{\infty} n A_n x^{(n-1)} = \sum_{n=1}^{\infty} \frac{1}{(n-1)!} \left( \frac{d^{(n-1)} B_z}{dx^{(n-1)}} \right)_0 x^{n-1} \quad (3)$$

so that

$$A_n = \frac{1}{n!} \left( \frac{d^{(n-1)} B_z}{dx^{(n-1)}} \right)_0 . \quad (4)$$

As a practical example of how one may identify the multipole components of a magnet by inspecting its symmetry, we digress a little to discuss the sextupole errors in the main dipoles of a large synchrotron.

Let us look at a simple dipole (Fig. 3). It is symmetric about the vertical axis and its field distribution will contain mainly even exponents of  $x$ , corresponding to odd  $n$  values: dipole, sextupole, decapole, etc. We can see, too, that cutting off the poles at a finite width can produce a virtual sextupole. Moreover, the remanent field pattern is frozen in at high field where the flux lines leading to the pole edges are shorter than those leading to the centre. The remanent magnetomotive force

$$\int H_c d\ell$$

is weaker at the pole edges, and the field tends to sag into a parabolic or sextupole configuration. This too produces a sextupole.

These three sources of sextupole error are the principle non-linearities in a large machine like the SPS. Note that these sextupole fields have no skew component. However, before launching into nonlinearities let us examine a simple linear resonance.

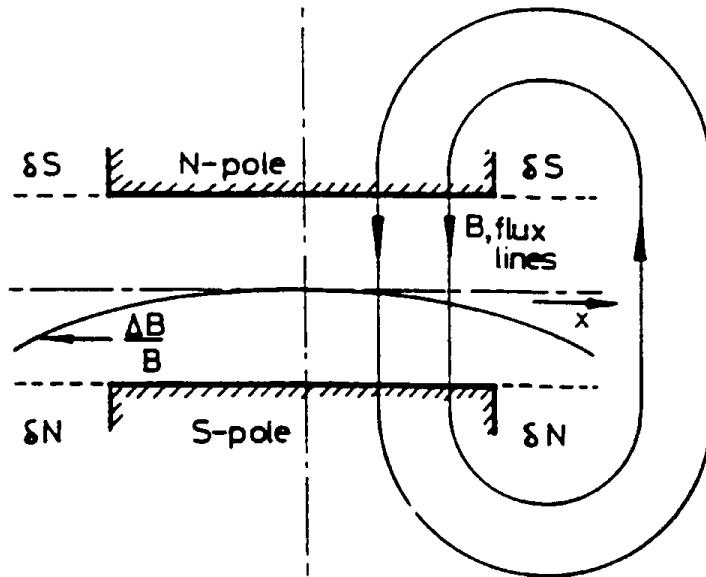


Fig. 3 The field in a simple dipole. The  $\delta N$  and  $\delta S$  poles superimposed on the magnet poles give the effect of cutting off the poles to a finite width.

### 3. SECOND-ORDER RESONANCE

A small elementary quadrupole of strength  $\delta(Kl)$  is located close to an  $F$  quadrupole where  $\beta_H = \hat{\beta}$ . Suppose a proton describes a circular trajectory of radius  $a = \sqrt{\epsilon\beta}$  and encounters the quadrupole at phase:

$$Q\phi(s) = Q\theta ,$$

where  $\theta$  is the azimuth which corresponds exactly to  $\phi$  at the quadrupoles of a FODO lattice.

The first step is to write down the unperturbed displacement at the small quadrupole:

$$x = a \cos Q\theta . \quad (5)$$

It receives a divergence kick (Fig. 4):

$$\Delta x' = \Delta(B\ell) / B\rho = \Delta(K\ell)x / B\rho . \quad (6)$$

The small change in  $\hat{\beta}\Delta x'$ ,

$$\Delta p = \hat{\beta}\Delta x' , \quad (7)$$

perturbs the amplitude,  $a$ , by

$$\Delta a = \Delta p \sin Q\theta .$$

Even more significantly there is a small phase advance (Fig. 4):

$$2\pi\Delta Q = \frac{\Delta p}{a} \cos Q\theta . \quad (8)$$

By successive substitution of Eqs. (7), (6). and (5), we get

$$2\pi\Delta Q = \hat{\beta} \frac{\Delta(\ell K)}{(B\rho)} \cos^2 Q\theta . \quad (9)$$

Over one turn the  $Q$  changes from the unperturbed  $Q$  by

$$\Delta Q = \frac{\hat{\beta}\Delta(\ell K)}{4\pi(B\rho)} (\cos 2Q\theta + 1) . \quad (10)$$

On the average this shifts  $Q$  by

$$\Delta Q = \frac{\hat{\beta}\Delta(\ell K)}{4\pi(B\rho)} . \quad (11)$$

Similarly the change in amplitude,  $a$ , is on average:

$$\frac{\Delta a}{a} \approx 2\pi\Delta Q .$$

The first term, however, tells us that, as the phase  $Q\theta$  on which the proton meets the quadrupole changes on each turn by  $2\pi \times$  (fractional part of  $Q$ ), the  $Q$ -value for each turn oscillates and may lie anywhere in a band

$$\delta Q = \frac{\hat{\beta}\Delta(\ell K)}{4\pi(B\rho)}$$

about the mean value.

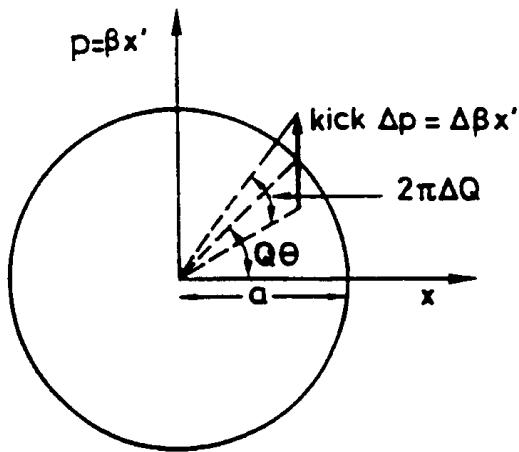


Fig. 4 Circle diagram shows effect of kick  $\delta p$  at phase  $Q\theta$  advancing phase by  $2\pi\Delta Q = (\Delta p \cos Q\theta)/a$

Suppose this band includes a half-integer  $Q$ -value. Eventually, on a particular turn, each proton will have exactly this half-integer  $Q$ -value ( $Q = p/2$ ).

Because the first term in Eq. (10) is  $\cos 2Q\theta$ , the amplitude increases by  $2\pi\Delta Q$  on the next and all subsequent turns. The proton has been perturbed by the  $\Delta(Kl)$  error to a  $Q$ -value where it "locks on" to a half-integer stopband. Once there, the proton repeats its motion every two turns, and the small amplitude increase from the perturbation  $\Delta a$  builds up coherently and extracts the beam from the machine.

We can visualize this in another way by saying that the half-integer line in the  $Q$  diagram,

$$2Q = p \quad (p = \text{integer}),$$

has a finite width  $\pm Q$  with respect to the unperturbed  $Q$  of the proton. Any proton whose unperturbed  $Q$  lies in this *stopband width* locks into resonance and is lost (Fig. 5).

In practice each quadrupole in the lattice of a real machine has a small field error. The  $\Delta(lK)$ 's are chosen from a random distribution with an r.m.s. value  $\Delta(lK)_{rms}$ . If the  $N$  focusing quadrupoles at  $\hat{\beta}$  have their principal effect, we can see that the r.m.s. expectation value for  $\delta Q$  is, from Eq. (11),

$$\langle \delta Q \rangle_{rms} = \sqrt{\frac{N}{2}} \frac{\hat{\beta} \Delta(K\ell)_{rms}}{4\pi B\rho} .$$

The factor  $\sqrt{2}$  comes from integrating over the random phase distribution. The statistical treatment is similar to that used for estimating closed orbit distortion.

Now let us use some Fourier analysis to see which particular azimuthal harmonic of the  $\delta(Kl)$  pattern drives the stopband. Working in normalized strength  $k = \Delta K/(B\rho)$  we analyse the function  $\delta(\beta k)$  into its Fourier harmonics with

$$\delta\beta k(s) = \sum \hat{\beta}_p k_p \cos(p\theta + \lambda) \quad (12)$$

and

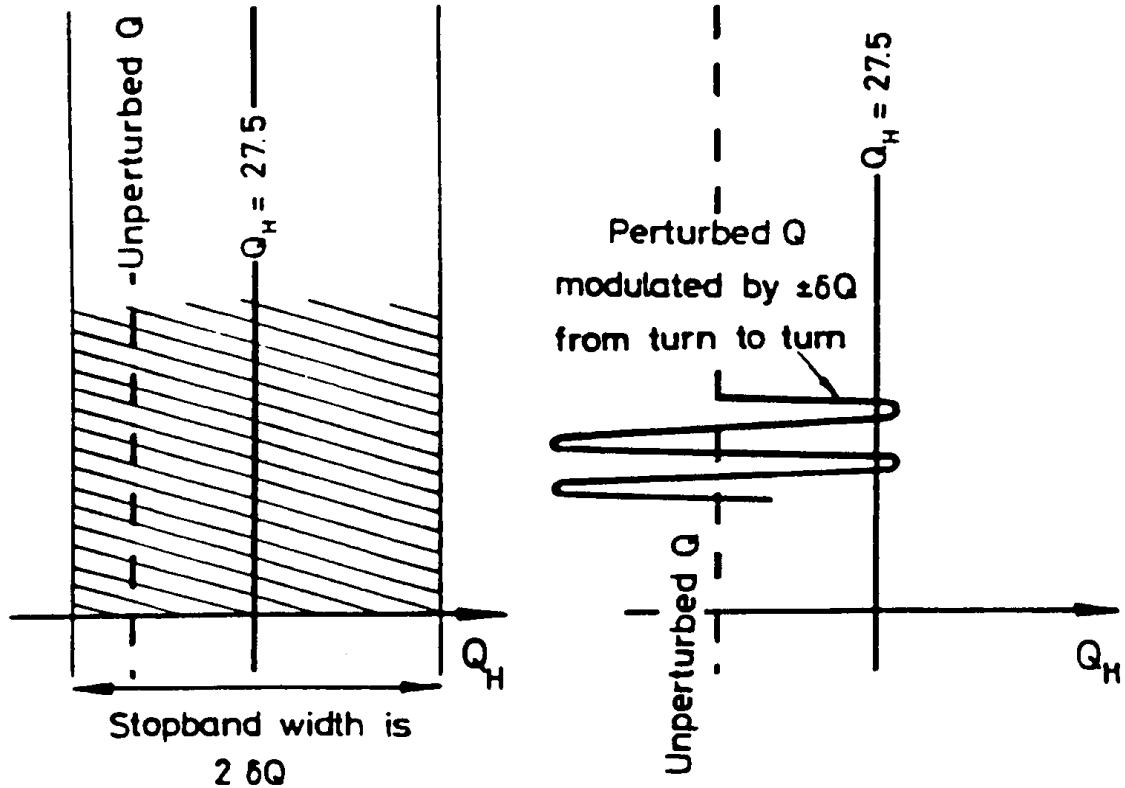


Fig. 5 Alternative diagrams showing perturbed  $Q$  and a stopband

$$\hat{\beta}k_p = \frac{1}{\pi R} \int_0^{2\pi} ds \delta[\beta k(s)] \cos(p\theta + \lambda) .$$

In general all harmonics i.e. all values of  $p$ , have equal expectation values in the random pattern of errors. We substitute the  $p^{\text{th}}$  term in Eq. (12) into Eq. (6) and work through the steps to obtain a new form for Eq. (10), namely:

$$2\pi\Delta Q = \int \frac{\hat{\beta}k_p}{2} \cos(p\theta + \lambda) \{\cos 2Q\theta + 1\} ds .$$

The integration can be simplified by writing  $ds = Rd\theta$ :

$$\Delta Q = \frac{\hat{\beta}k_p R}{4\pi} \int_0^{2\pi} \cos 2Q\theta \cos(p\theta + \lambda) d\theta .$$

The integral is only finite over many betatron oscillations when the resonant condition is fulfilled:

$$2Q = p .$$

We have revealed the link between the azimuthal frequency  $p$  in the pattern of quadrupole errors and the  $2Q = p$  condition which describes the stopband. For example, close to  $Q = 27.6$  in the SPS lies the half-integer stopband  $2Q = 55$ . The azimuthal Fourier component which

drives this is  $p = 55$ . Similarly, a pattern of correction quadrupoles, powered in a pattern of currents which follows the function

$$i = i_0 \sin(55\theta + \lambda) ,$$

can be used to compensate the stopband by matching  $i_0$  and  $\lambda$  empirically to the amplitude of the driving term in the error pattern.

This has been used at the SPS, and in other machines powering sets of harmonic correction quadrupoles, each with its own power supply. We look for a sudden beam loss due to a strong stopband at some point in the cycle where  $Q'$  and  $\Delta p/p$  are large and gradient errors important. This loss will appear as a downward step in the beam current transformer signal. We then deliberately make  $Q$  sit on the stopband at that point to enhance the step and alter the phase and amplitude of the azimuthal current patterns of the harmonic correctors to minimize the loss. We may have to do this at various points in the cycle with different phase and amplitude.

Two sets of such quadrupoles are used: one set near  $F$  lattice quadrupoles affecting mainly  $2Q_H = 55$ ; the other set near  $D$  quadrupoles affecting  $2Q_V = 55$ .

#### 4. THE THIRD-INTEGER RESONANCE

The third-integer stopbands are driven by sextupole field errors and are therefore non-linear. First imagine a single short sextupole of length 1, near a horizontal maximum beta location. Its field is

$$\Delta B = \frac{d^2 B_z}{dx^2} x^2 = \frac{B''}{2} x^2 , \quad (13)$$

and it kicks a particle with betatron phase  $Q\theta$  by

$$\Delta p = \frac{\beta \ell B''}{2B\rho} x^2 = \frac{\beta \ell B'' a^2}{2B\rho} \cos^2 Q\theta \quad (14)$$

inducing increments in phase and amplitude,

$$\frac{\Delta a}{a} = \frac{\Delta p}{a} \sin Q\theta = \frac{\beta \ell B'' a}{2B\rho} \cos^2 Q\theta \sin Q\theta \quad (15)$$

$$\Delta \phi = \frac{\Delta p}{a} \cos Q\theta = \frac{\beta \ell B'' a}{2B\rho} \cos^3 Q\theta \quad (16)$$

$$= \frac{\beta \ell B'' a}{8B\rho} (\cos 3Q\theta + 3 \cos Q\theta) . \quad (17)$$

Suppose  $Q$  is close to a third integer, then the kicks on three successive turns appear as in Fig. 6. The second term in Eq. (17) averages to zero over three turns and we are left with a phase shift:

$$2\pi\Delta Q = \Delta \phi = \frac{\beta \ell B'' a \cos 3Q\theta}{8B\rho} . \quad (18)$$

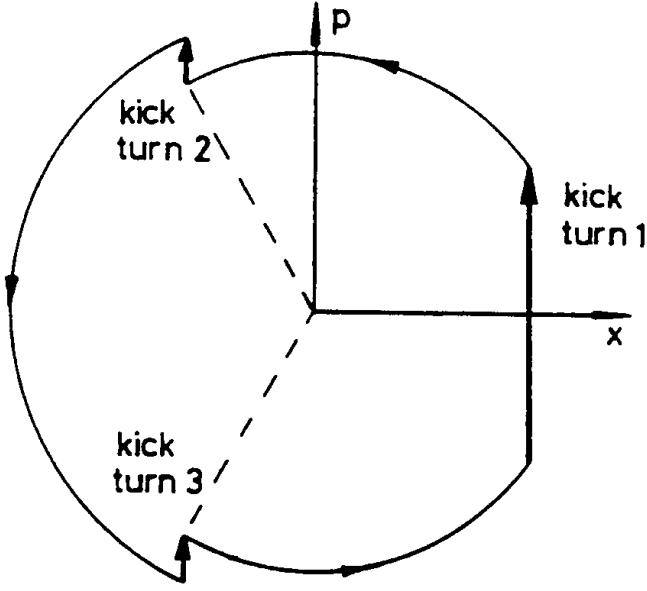


Fig. 6 Phase-space trajectory on a 3rd-order resonance

We can again guess how resonances arise. Close to  $Q = p/3$ , where  $p$  is an integer,  $\cos 3Q\theta$  varies slowly, wandering within a band about the unperturbed  $Q_0$  as in Fig. 5:

$$Q_0 - \frac{\beta \ell B'' a}{16\pi B\rho} < Q < Q_0 + \frac{\beta \ell B'' a}{16\pi B\rho}. \quad (19)$$

As in the case of the half-integer resonance this is the stopband width but in reality is a perturbation in the motion of the particle itself.

We can write the expression for amplitude perturbation

$$\frac{\Delta a}{a} = \frac{\beta \ell B'' a}{8B\rho} \sin 3Q\theta. \quad (20)$$

Suppose the third integer  $Q$ -value is somewhere in the band. Then, after a sufficient number of turns, the perturbed  $Q$  of the machine will be modulated to coincide with  $3p$ . On each subsequent revolution this increment in amplitude builds up until the particle is lost. Growth is rapid and the modulation of  $Q$  away from the resonant line is comparatively slow.

Looking back at the expressions, we find that the resonant condition,  $3Q = \text{integer}$ , arises because of the  $\cos^3 Q\theta$  term in Eq. (16), which in turn stems from the  $x^2$  dependence of the sextupole field. This reveals the link between the order of the multipole and that of the resonance.

We see that the  $a^2$  in Eq. (14) leads to a linear dependence of width upon amplitude. This term was  $a^1$  in the case of the half integer resonance which led to a width which was independent of amplitude and will become  $a^3$  in the case of a fourth-order resonance giving a parabolic dependence of width upon amplitude.

It is also worth noting that the second term in Eq. (17), which we can ignore when away from an integer  $Q$ -value, suggests that sextupoles can drive integer stopbands as well as third

integers. Inspection of the expansion of  $\cos^n\theta$  will suggest the resonances which other multipoles are capable of driving.

Returning to the third-order stopbands, we note that both stopband width and growth rate are amplitude-dependent. If  $Q_0$  is a distance  $\Delta Q$  from the third integer resonance, particles with amplitudes less than

$$a < \frac{16\pi(B\rho)\Delta Q}{\beta\ell B''} \quad (21)$$

will never reach a one third integer  $Q$  and are in a central region of stability. Replacing the inequality by an equality, we obtain the amplitude of the metastable fixed points in phase space where there is resonant condition but infinitely slow growth (Fig. 7).

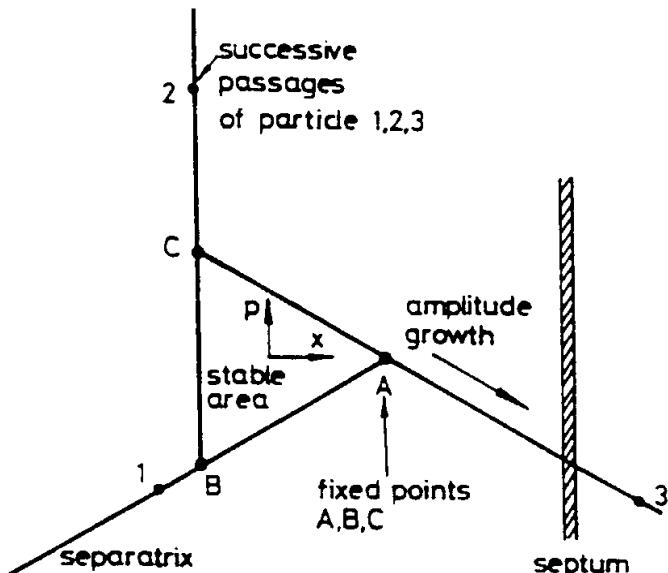


Fig. 7 Third-order separatrix

The symmetry of the circle diagram suggests there are three fixed points at  $\theta = 0, 2\pi/3$ , and  $4\pi/3$ . For a resonance of order,  $n$ , there will be  $n$  such points.

The fixed points are joined by a separatrix, which is the bound of stable motion. A more rigorous theory, which takes into account the perturbation in amplitude, would tell us that the separatrix is triangular in shape with three arms to which particles cling on their way out of the machine.

We have seen how a single sextupole can drive the resonance. Suppose now we have an azimuthal distribution which can be expressed as a Fourier series:

$$B''(\theta) = \sum B_p'' \cos p\theta . \quad (22)$$

Then

$$\Delta\phi = \sum_p \int \frac{\beta B_p''}{8B\rho} \cos 3Q\theta \cos p\theta d\theta . \quad (23)$$

This integral is large and finite if

$$p = 3Q . \quad (24)$$

As in the earlier case of the second-order resonances this reveals why it is a particular harmonic in the azimuthal distribution which drives the stopband. It is not just the Fourier spectrum of  $B''(\theta)$  but of  $\beta B''(\theta)$  which is important in this context. Periodicities in the lattice and in the multipole pattern can thus mix to drive resonances.

This is particularly important since some multipole fields, like the remanent field pattern of dipole magnets, are inevitably distributed in a systematic pattern around the ring. This pattern is rich in the harmonics of  $S$ , the superperiodicity. Even when this is not the case and errors are evenly distributed, any modulation of beta which follows the pattern of insertions can give rise to systematic driving terms. It is an excellent working rule to keep any systematic resonance, i.e.

$$\ell Q_H + m Q_V = S \text{ (superperiod number)} \times \text{integer} = p \quad (25)$$

out of the half integer square in which  $Q$  is situated. This is often not easy in practice.

As in the second-order case, the third-order stopbands can be compensated with sets of multipoles powered individually to generate a particular Fourier component in their azimuthal distributions. The above equation defines four numerical relations between  $Q_H$  and  $Q_V$  which are resonant. The keen student can verify this with an extension to the mathematics of the previous section. He will find that two of the lines are sensitive to errors of a sextupole configuration with poles at the top and bottom, the other two to sextupoles with poles symmetrical about the median plane (Fig. 2). By permuting these two kinds of sextupoles with the two types of location, we can attack the four lines more or less orthogonally.

## 5. GENERAL NUMEROLOGY OF RESONANCES

We have seen how the  $Q$ -value at which the resonance occurs is directly related to a frequency in the azimuthal pattern of variation of multipole strength. We can now generalise this.

Suppose the azimuthal pattern of a multipole of order  $n$  can be Fourier analysed:

$$B^{(n-1)}(\theta) = \sum_p B_p^{(n-1)} \cos p\theta, \quad (26)$$

where  $\theta$  is an azimuthal variable, range 0 to  $2\pi$ . We shall show that if the resonance is in one plane only, a particular component,  $p = nQ$ , of this Fourier series, drives it. For example, the 83rd azimuthal harmonic of sextupole ( $n = 3$ ) drives the third-order resonance at  $Q = 27.66$ . The more general expression is

$$\ell Q_H + m Q_V = p \quad (27)$$

$$|\ell| + |m| = n \text{ (an integer).} \quad (28)$$

Each  $n$ -value defines a set of lines in Fig. 1, four for third-order resonances, five for fourth-order, etc. Each line corresponds to a different homogeneous term in the multipole Cartesian expansion (Table 1). Some are excited by regular multipoles, others by skew multipoles.

## 6. SLOW EXTRACTION USING THE THIRD-ORDER RESONANCE

So far we have thought of resonances as a disease to be avoided, yet there is at least one useful function that they can perform.

We have seen that a third-order stopband extracts particles above a certain amplitude, the amplitude of the unstable fixed points which define a separatrix between stability and instability

(Fig. 7). The dimensions of the separatrix, characterized by  $a$  are determined by  $\Delta Q$ , the difference between the unperturbed  $Q$  and the stopband. As one approaches the third integer by, say, increasing the focusing strength of the lattice quadrupoles,  $\Delta Q$  shrinks, the unstable amplitude,  $a$ , becomes smaller and particles are squeezed out along the three arms of the separatrix. If we make  $\Delta Q$  shrink to zero over a period of a few hundred milliseconds, we can produce a rather slow spill extraction.

At first sight we might expect only one third of the particles to migrate to positive  $x$ -values since there are three separatrices, but it should be remembered that a particle jumps from one arm to the next each turn, finally jumping the extraction septum on the turn when its displacement is largest. The septum is a thin walled deflecting magnet at the edge of the aperture.

The growth increases rapidly as particles progress along the unstable separatrix, and if the stable area is small compared with the distance between beam and septum, the probability of a particle striking the septum rather than jumping over it is small. It clearly helps to have a thin septum. The SPS it is a comb of wires forming a plate of an electrostatic deflector.

Magnet or quadrupole ripple can cause an uneven spill, making the  $Q$  approach the third integer in a series of jerks thus modulating the rate at which particles emerge. A spread in momentum amongst the particles can help, however, since if the chromaticity is finite, we will have swept through a larger range of  $Q$ -values before all separatrices for all momenta have shrunk to zero. The larger  $Q$  change reduces the sensitivity to magnet ripple.

## 7. LANDAU DAMPING WITH OCTUPOLES

Another beneficial effect of multipoles is the use of octupoles to damp coherent transverse instabilities due to the beam's own electromagnetic field.

For a transverse instability to be dangerous, the growth time must win over other mechanisms which tend to destroy the coherent pattern and damp out the motion. One such damping mechanism is the  $Q$ -spread in the beam. Coherent oscillations decay, or become dephased, in a number of betatron oscillations comparable to  $1/\Delta Q$ , where  $\Delta Q$  is the  $Q$ -spread in the beam. This corresponds to a damping time, expressed in terms of the revolution frequency,  $\omega_0/2\pi$ :

$$\tau_d = \frac{2\pi}{\omega_0 \Delta Q} , \quad (29)$$

which is just the inverse of the spread in frequencies of the oscillators involved, i.e. the protons. The threshold for the growth of the instability is exceeded when  $\tau_g$  (which increases with intensity) exceeds  $\tau_d$

$$\tau_g = \frac{2\pi}{\omega_0 \Delta f} . \quad (30)$$

This is a very general argument which affects all instability problems involving oscillators and is an example of *Landau damping*. Thinking of it another way, we can say that the instability never gets a chance to grow if the oscillators cannot be persuaded to act collectively for a time  $\tau_g$ . If they have a frequency spread  $\Delta f$ , the time for which they can act concerted is just  $1/\Delta f$ .

Unfortunately, in our quest for a small  $\Delta Q$  to avoid lines in the  $Q$  diagram by correcting chromaticity, improvements in single particle dynamics can lower the threshold intensity for the instability. A pure machine is infinitely unstable. In practice, at the SPS this happens at about  $5 \times 10^{12}$  particles per pulse if  $\Delta Q$  is less than 0.02 and  $\tau_g$  about 1 msec. Suddenly the beam begins to snake under the influence of the resistive wall instability. A large fraction of the beam is lost before stability is restored.

The first remedy is to increase  $\Delta Q$ . Landau damping octupoles are installed for this purpose in the SPS. Octupoles produce an amplitude  $Q$ -dependence which is thought to be more effective than the momentum-dependent  $Q$ -spread produced by sextupoles. Each particle changes in momentum during a synchrotron oscillation, and in a time comparable to  $\tau_g$  all particles have the same mean momentum. Sextupoles do not spread the mean  $Q$  of the particles. Octupoles, producing an amplitude  $Q$ -dependence, do.

The circle diagram can be used to calculate the effect of an octupole which gives a kick:

$$\Delta p = \beta \frac{\Delta(B\ell)}{B\rho} = \frac{\beta\ell B'''}{3!(B\rho)} a^3 \cos^3 Q\theta . \quad (31)$$

The change in phase is

$$2\pi\Delta p = \Delta\phi = \frac{\beta\ell B''' a^2 \cos^4 Q\theta}{6(B\rho)} , \quad (32)$$

which averages to

$$\Delta Q = \frac{\beta\ell B''' \beta a^2}{32\pi B\rho} . \quad (33)$$

Of course if the octupoles are placed around the ring they can excite fourth-order resonances. A good rule is to have as many of them as possible and to distribute them at equal intervals of betatron phase. If there are  $S$  octupoles thus distributed their Fourier harmonics are  $S, 2S$ , etc. and they can only excite structure resonances near  $Q$  values:

$$4Q = S \times \text{an integer} .$$

Although these systematic resonances are very strong it should not be difficult to choose  $S$  so that  $Q$  is not in the same integer square as one of the values of  $nS/4$ .

\* \* \*

## FURTHER READING

G. Guignard, Effets des Champs Magnétiques Perturbateurs d'un Synchrotron - CERN 70-24 (1970), p. 67-105.

M.H. Blewett (Editor), Theoretical Aspects of the Behaviour of Beams in Accelerators and Storage Rings - CERN 77-13 (1977), p. 111-138.

G. Guignard, A General Treatment of Resonances in Accelerators, CERN 78-11 (1978).

# DYNAMICS AND ACCELERATION IN LINEAR STRUCTURES

*J. Le Duff*

Laboratoire de l'Accélérateur Linéaire  
Bat. 200, Centre d'Orsay, 91405 Orsay, France

## 1. BASIC METHODS OF LINEAR ACCELERATION

### 1.1 Early days

In principle a linear accelerator is one in which the particles are accelerated on a linear path. Then the most simple scheme is the one which uses an electrostatic field as shown in Fig. 1. A high voltage is shared between a set of electrodes creating an electric accelerating field between them. The disadvantage of such a scheme, as far as high energies are concerned, is that all the partial accelerating voltages add up at some point and that the generation of such high electrostatic voltages will be rapidly limited (a few ten MV). This type of accelerator is however currently used for low energy ion acceleration, and is better known as the Van De Graaf accelerator.

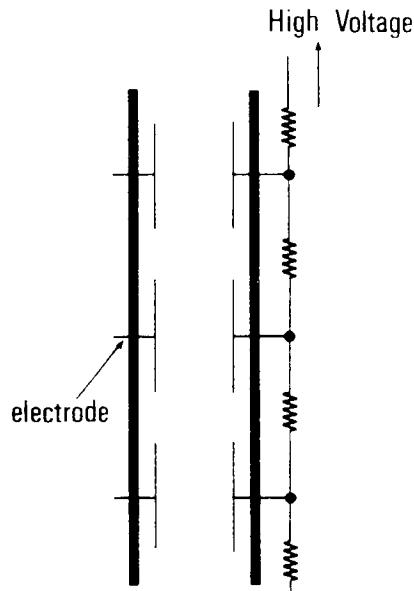


Fig. 1 Electrostatic accelerator scheme

In the late 1920's propositions were made, essentially by R. Wideroe, to avoid the limitation of electrostatic devices due to voltage superposition. The proposed scheme, later on (early 1930's) improved by E. Lawrence and D. Sloan at the Berkeley University, is shown on Fig. 2.

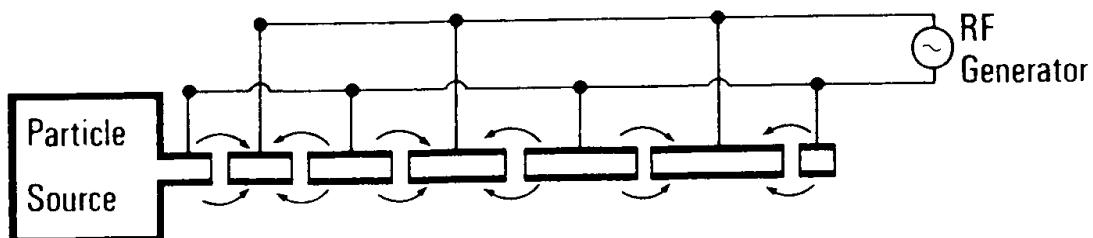


Fig. 2 Wideroe-type accelerator

An oscillator (7 MHz at that time) feeds alternately a series of drift tubes in such a way that particles see no field when travelling inside these tubes while they are accelerated in between. The last statement is true if the drift tube length  $L$  satisfies the synchronism condition:

$$L = \frac{vT}{2}$$

where  $v$  is the particle velocity ( $\beta c$ ) and  $T$  the period of the a.c. field. This scheme does not allow continuous acceleration of beams of particles.

## 1.2 Improved methods for non-relativistic particles

Consider a proton of 1 MeV kinetic energy entering the previous structure. At a frequency of 7 MHz such a particle, with  $\beta = v/c = 4.6 \cdot 10^{-2}$ , will travel a distance of roughly 1 meter in half a cycle. Clearly the length of the drift tubes will soon become prohibitive at higher energies unless the input RF frequency is increased.

Higher-frequency power generators only became available after the second world war, as a consequence of radar developments.

However at higher frequencies the system, which is almost capacitive, will radiate a large amount of energy; as a matter of fact if one considers the end faces of the drift tubes as the plates of a capacitor, the displacement current flowing through it is given by

$$I = \omega CV$$

where  $C$  is the capacitance between the drift tubes,  $V$  the accelerating voltage and  $\omega$  the angular frequency in use. It is therefore convenient to enclose the gap existing between drift tubes in a cavity which holds the electromagnetic energy in the form of a magnetic field (inductive load) and to make the resonant frequency of the cavity equal to that of the accelerating field (Fig. 3). In that case the accelerator would consist of a series of such cavities fed individually with power sources.

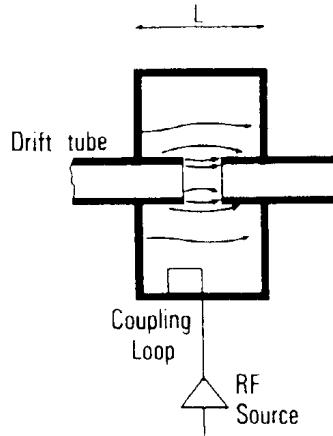


Fig. 3 Single-gap accelerating structure

Such single-gap cavities could also be placed adjacent to each other as shown on Fig. 4. In the  $2\pi$  mode case, since the resulting wall current is zero, the common walls between cavities become useless. Then a variant of that scheme consists of placing the drift tubes in a single resonant tank such that the field has the same phase in all gaps. Such a resonant accelerating structure was invented by L. Alvarez in 1945 and was followed by the construction of a 32 MeV proton drift tube linac (Fig. 5) powered by 200 MHz war surplus radar equipment.

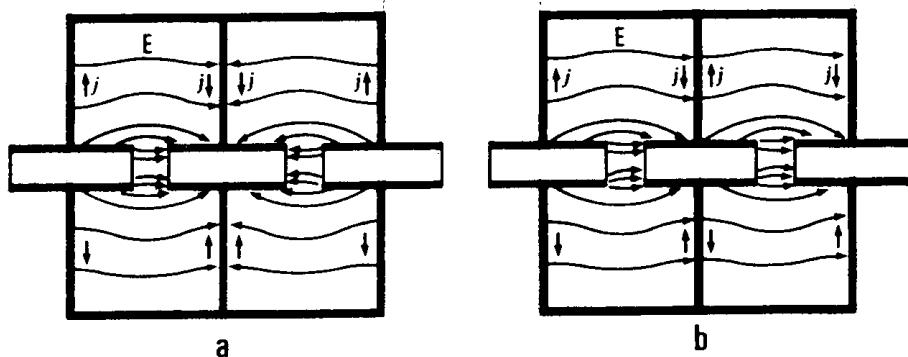


Fig. 4 Adjacent single-gap cavities: a)  $\pi$  mode, b)  $2\pi$  mode

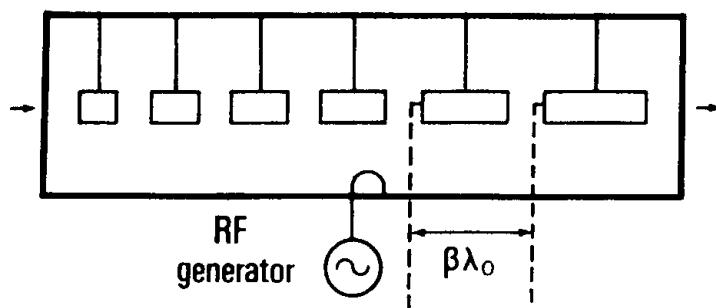


Fig. 5 Alvarez-type structure

In the  $2\pi$  mode of operation the synchronism condition is:

$$L = vT = \beta\lambda_0$$

where  $\lambda_0$  is the free space wavelength at the operating frequency. Notice that in Fig. 5 the drift tubes are maintained by metallic rods to the tank walls.

The Alvarez structure is still used for protons, as well as heavy ions, operating mostly at 200 MHz. Most of our present day proton linear accelerators are used as injectors for circular machines such as synchrotrons and their energy lies from 50 MeV to 200 MeV. At 200 MeV protons are still weakly relativistic with  $\beta = 0.566$ .

*Note:* Since the progress in methods of acceleration came from the use of resonant structures which can provide high accelerating field with less power consumption, the new definition of a linear accelerator or "Linac" implied machines in which particles are accelerated on a linear path by radio frequency fields. Then electrostatic devices no more appear in this definition, but it is worthwhile mentioning that they are used as front-end proton linacs.

### 1.3 The case of ultra-relativistic particles

While  $\beta$  is getting close to unity for protons of 10 GeV kinetic energy,  $\beta$  is almost unity for electrons of 10 MeV. Hence above these energies the particles will have a constant velocity  $v = c$  and the length of the drift tubes will remain constant as well. The higher velocity needs higher frequencies. However triode and tetrode tubes could not handle high RF power at high frequency. The invention of the klystron in 1937 and its successful development during the war led to high power sources at 3000 MHz. At this frequency the free-space wavelength is 10 cm, small enough that the perspective of accelerating electrons to high energies soon became an aim.

At the same time emerged the idea that ultrarelativistic particles could be accelerated by travelling guided waves. It is a matter of fact that in a resonant structure the standing wave pattern can be expanded into two travelling waves, one which travels in synchronism with the particle and the backward wave which has no mean effect on the particle energy.

However TM modes (with an electric field in the direction of propagation) in rectangular or cylindrical guides have phase velocities bigger than  $c$ . Then it was necessary to bring the phase velocity at the level of the particle velocity ( $v_p \sim c$ ) and to do so the simplest method consists of loading the structure with disks as shown on Fig. 6, where the size of the holes determines the degree of coupling and so determines the relative phase shift from one cavity to the next. When the dimensions ( $2a$ ,  $2b$ ) have been tailored correctly the phase changes from cavity to cavity along the accelerator to give an overall phase velocity corresponding to the particle velocity.

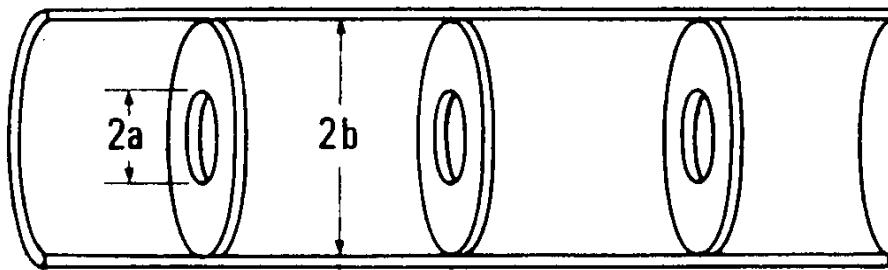


Fig. 6 Disk-loaded structure

This type of structure will continuously accelerate particles as compare to the drift tube structure which gives a discontinuous acceleration corresponding to the successive gaps.

Figure 7 is a more complete drawing of such a travelling-wave structure showing both, the input coupler which matches the source to the structure and the output coupler which matches the structure to an external load (resistive load for instance) to avoid the backward wave.

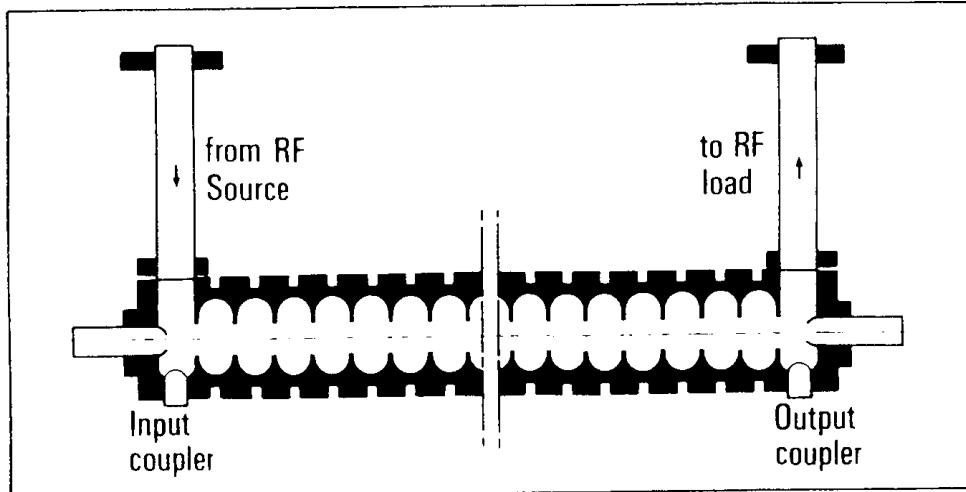


Fig. 7 Travelling-wave accelerating structure

These structures generally operate in the  $\pi/2$  mode or the  $2\pi/3$  mode. For the former the height of each cell is equal to  $\lambda/4$  while it is equal to  $\lambda/3$  for the latter. This is important, as will be seen later, for the electromagnetic energy to propagate. The interesting thing with travelling-wave structures, in which the energy propagates relatively fast, is that the RF power source can

be pulsed during a short period corresponding to the filling time of the structure. In this pulsed mode of operation much higher peak power pulses can feed the structure, increasing the accelerating field. As a consequence only pulsed beams can be accelerated leading to small duty cycles.

Standing-wave structures can also be used for ultrarelativistic particles. In that case the  $\pi$  mode of operation is efficient, where the field has opposite phase in two adjacent cells. This type of structure as shown on Fig. 8, often called "nose cone structure", is very similar to the drift tube one in which the length of the tubes has been made very small. A variant of this scheme is used in the high energy proton linac ( $E = 800$  MeV) at Los Alamos, where the coupling between cavities has been improved by adding side coupled resonant cavities as sketched on Fig. 9.

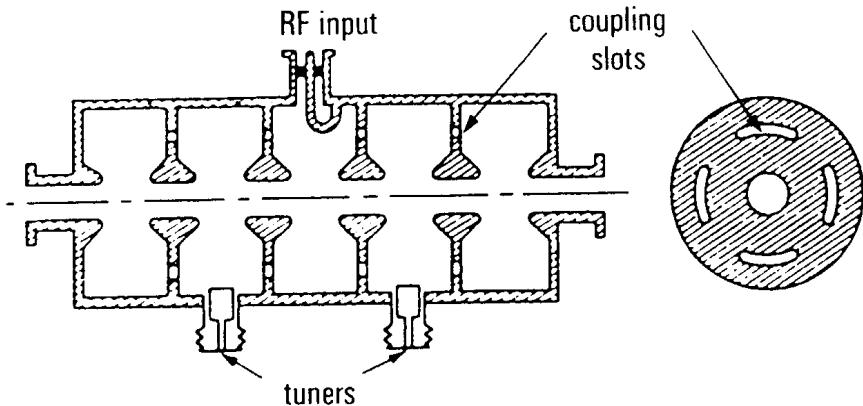


Fig. 8 Nose-cone structure

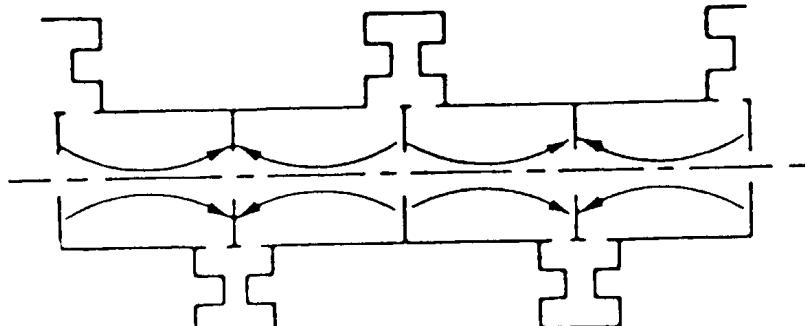


Fig. 9 Side-coupled structure

#### 1.4 Induction linac

Resonant structures as described previously cannot handle very high beam currents. The reason is that the beam induces a voltage proportional to the circulating current and with a phase opposite to that of the RF accelerating voltage. This effect known as "beam loading" disturbs the beam characteristics and can even destroy the beam by some instability mechanism.

A cure for such an effect in the case of very high currents consists of producing an accelerating field with a very low Q resonator. This is obtained with an induction accelerator module (Fig. 10) in which a pulsed magnetic field produces an electric field component, according to Maxwell equations, just similar to the betatron principle.

The accelerator will consist of an array of such modules triggered at a rate compatible with the particle velocity, and fed by high power short pulse generators.

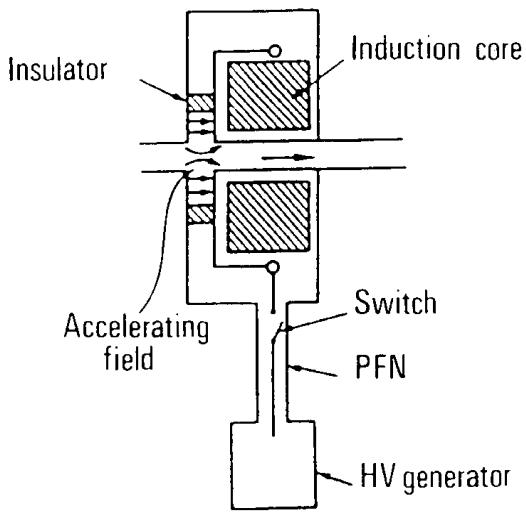


Fig. 10 Linear induction accelerator module

### 1.5 Radio frequency quadrupole (RFQ)

At quite low  $\beta$  values (for example low energy protons) it is hard to maintain high currents due to the space charge forces of the beam which have a defocusing effect.

In 1970 I.M. Kapchinski and V.A. Teplyakov from the Soviet Union proposed a device in which the RF fields which are used for acceleration can serve as well for transverse focusing. The schematic drawing of an RFQ is shown on Fig. 11. The vanes which have a quadrupole symmetry in the transverse plane have a sinusoidal shape variation in the longitudinal direction. In recent years these devices have been built successfully in many laboratories making it possible to lower the gun accelerating voltage for protons and heavy ions to less than 100 kV as compared to voltages above 500 kV which could only be produced earlier by large Cockcroft-Walton electrostatic generators.

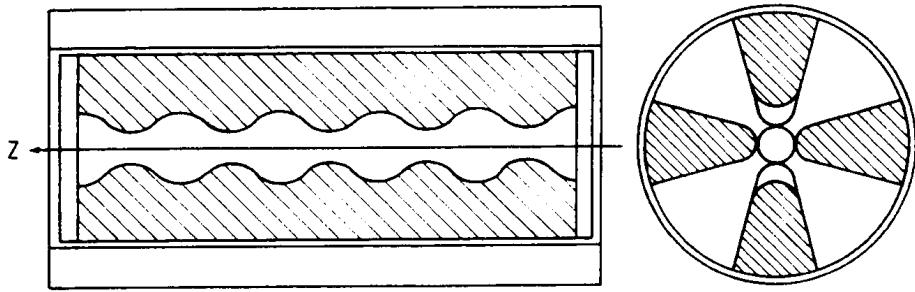


Fig. 11 Schematic drawing of an RFQ resonator

### 1.6 Other methods and future prospects

Among the other methods of acceleration one can at least distinguish between two classes: collective accelerators and laser accelerators. In both cases the idea is to reach much higher gradients in order to produce higher energies keeping the overall length of the accelerator at a reasonable level.

Collective accelerators are already in use for ion acceleration but up to now they never reached the desirable high gradients. The oldest idea of collective acceleration is the Electron Ring Accelerator (ERA) where an intense electron beam of compact size is produced in a compressor (Fig. 12). The electron ring is then accelerated either by an electric field or by a

pulse magnetic field (induction acceleration) and loaded with ions. Through the space charge effect the electrons (hollow beam) will take the ions along.

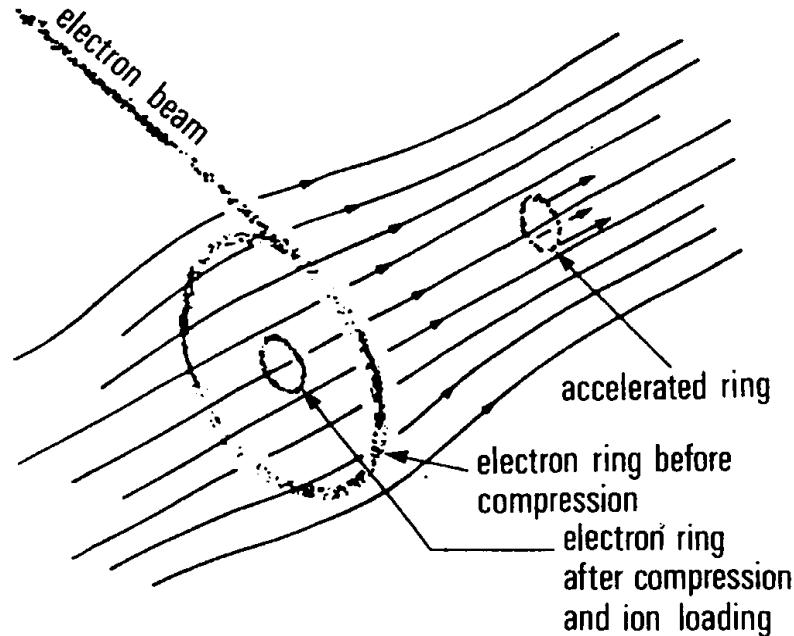


Fig. 12 Principle of the Electron Ring Accelerator (ERA)

Laser accelerators hold out the promise of reaching high energies with a technology which is new to accelerator physicists. Plasma media can be used to lower the velocity of the laser wave.

It is also worthwhile to mention that extensions of conventional techniques are also studied extensively for very high energy electron linacs.

## 2. FUNDAMENTAL PARAMETERS OF ACCELERATING STRUCTURES

### 2.1 Transit time factor

Consider a series of accelerating gaps as in the Alvarez structure (Fig. 13a) and assume the corresponding field in the gap to be independent of the longitudinal coordinate  $z$  (Fig. 13 b). If  $V$  is the maximum voltage in the gap, the accelerating field is:

$$E_z = \frac{V}{g} \cos \omega t$$

If the particle passes through the center of the gap at  $t = 0$  with a velocity  $v$ , its coordinate is:

$$z = vt$$

and its total energy gain is:

$$\begin{aligned} \Delta E &= \int_{-g/2}^{+g/2} \frac{eV}{g} \cos \omega \frac{z}{v} dz \\ &= eV \frac{\sin \theta / 2}{\theta / 2} = eVT \end{aligned}$$

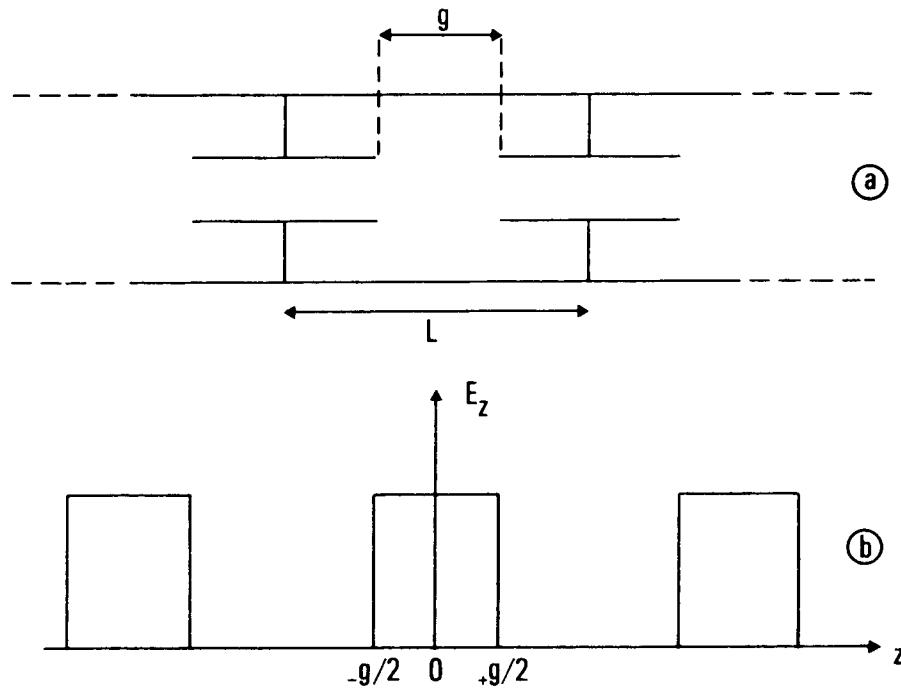


Fig. 13 Approximate field pattern in a drift tube accelerator

where

$$\theta = \frac{\omega g}{v}$$

is called the transit angle and  $T$  is the transit-time factor:

$$T = \frac{\sin \theta / 2}{\theta / 2}$$

For a standing-wave structure operating in the  $2\pi$  mode and where the gap length is equal to the drift tube length:

$$g = \beta \lambda_o / 2$$

one gets:

$$T = 0.637 .$$

To improve upon this situation, for a given  $V$  it is advantageous to reduce the gap length  $g$  which leads to larger drift tubes as in the Alvarez design. However a too large reduction in  $g$  will lead to sparking, for a given input power per meter, due to an excessive local field gradient. Usual values of  $T$  lie around 0.8. In the more general case where the instantaneous field is not homogeneous through the gap, the transit-time factor is given by:

$$T = \frac{\left| \int E_z(z) e^{j\omega t} dz \right|}{\int E_z(z) dz}$$

The transit time factor generally shows the amount of energy which is not gained due to the fact that the particle travels with a finite velocity in an electric field which has a sinusoidal time variation. However this factor may become meaningless, for instance if the mode is such

that the denominator is equal to zero while the numerator remains finite as would be the case for a TM<sub>011</sub> mode in a pill-box cavity (see Fig. 14). So one has to be careful when using this concept.

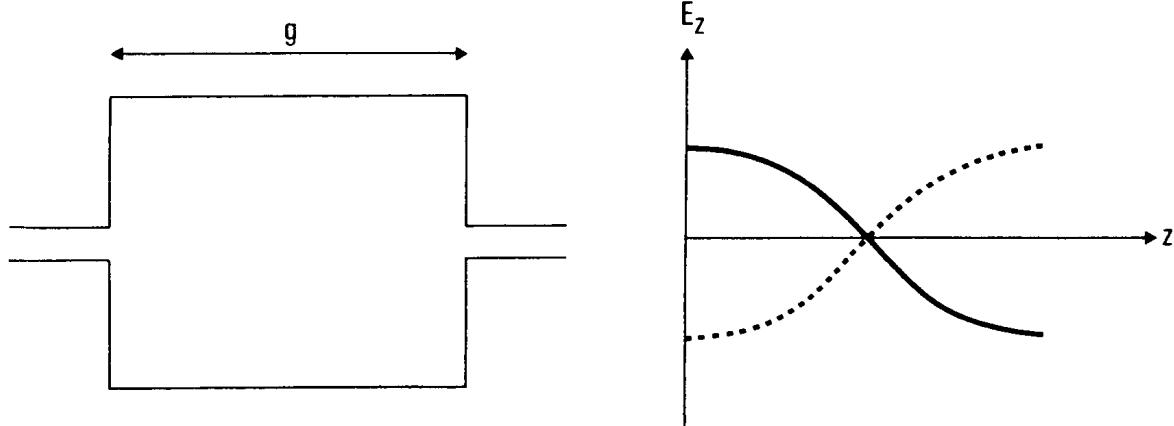


Fig. 14 TM<sub>011</sub> mode in a pill-box cavity

Exercise: Energy gain when the field  $E_z$  in the gap varies with  $z$

One has:

$$\Delta E = e \Re e \int_0^g E_z(z) e^{j\omega t} dz$$

with

$$\omega t = \omega \frac{z}{v} - \psi_p$$

where  $\psi_p$  is the phase of the particle, relative to the RF, when entering the gap. Hence

$$\begin{aligned} \Delta E &= e \Re e \left[ e^{-j\psi_p} \int_0^g E_z(z) e^{j\omega \frac{z}{v}} dz \right] \\ &= e \Re e \left[ e^{-j\psi_p} e^{j\psi_i} \left| \int_0^g E_z(z) e^{j\omega \frac{z}{v}} dz \right| \right] \end{aligned}$$

By introducing  $\phi = \psi_p - \psi_i$  one finally gets:

$$\Delta W = e \left| \int_0^g E_z(z) e^{j\omega \frac{z}{v}} dz \right| \cos \phi$$

which has a maximum value for  $\phi = 0$ .

Now  $\phi$  appears as the phase of the particle referred to the particular phase which would yield the maximum energy.

## 2.2 Shunt impedance

The shunt impedance  $R_s$  for an RF cavity operating in the standing wave mode is a figure of merit which relates the accelerating voltage  $V$  to the power  $P_d$  dissipated in the cavity walls:

$$P_d = \frac{V^2}{R_s} .$$

The shunt impedance is very often defined as a quantity per unit length. So, a more general definition which takes also care of travelling-wave structures is:

$$\frac{dP}{dz} = -\frac{E_z^2}{r} \quad \text{with } r = \frac{R_s}{L}$$

where  $L$  is the cavity length,  $r$  the shunt impedance per unit length,  $E_z$  the amplitude of the accelerating field, and  $\frac{dP}{dz}$  the fraction of the input power lost per unit length in the walls (another fraction will go into the beam). The sign in the right hand side means that the power flowing along a travelling-wave structure decreases due to the losses.

In the case of standing-wave cavities an uncorrected shunt impedance  $Z$  is sometimes defined (computer codes for designing cavities) where  $V$  is the integral of the field envelope along the gap. Then, to take care of the transit time factor the true shunt impedance becomes

$$R_s = Z T^2 .$$

Shunt impedances up to 35 MΩ/m are reached in proton linacs operating at 200 MHz and relatively low energy, while shunt impedances up to 100 MΩ/m can be obtained at 3 GHz in electron linacs. For the latter a peak power of 50 MW (for instance supplied by a high power pulsed klystron) would give an accelerating gradient of 70 MV/m in a 1 meter-long structure.

However, most of the present electron linacs work in the range of 10 to 20 MV/m with less efficient structures and lower peak power from more conventional pulsed klystrons.

If a standing-wave structure, with shunt impedance  $R_s$ , is used in the travelling-wave mode then the shunt impedance is doubled. This comes from the fact that a standing wave can be considered as the superposition of two travelling waves of opposite direction, each wave leading to power losses in the walls.

It is desirable to have a shunt impedance per unit length  $r$  as high as possible. Let's have a look to the dependance of  $r$  upon the operating frequency:

- the RF power loss per unit length is proportional to the product of the square of the wall current  $i_w$  and the wall resistance  $r_w$  per unit length:

$$\frac{dP}{dz} \propto i_w^2 r_w$$

- the axial electric field  $E_z$  is proportional to the wall current divided by the radius  $b$  of the cavity:

$$E_z \propto i_w / b$$

- the wall resistance  $r_w$  per unit length is equal to the resistivity  $\rho$  of the wall material divided by the area of the surface through which the current is flowing:

$$r_w = \rho / 2\pi b \delta$$

where  $\delta$  is the skin depth given by:

$$\delta = (2\rho / \omega\mu)^{1/2}$$

and  $\mu$  is the permeability of the walls. Combining all these expressions and knowing that  $b \propto 1/\omega$  yields the result:

$$r \propto \sqrt{\omega}$$

which shows, from the viewpoint of RF power economy, that it is better to operate at higher frequencies. But there is however a limit in going to very high frequencies due to the fact that the aperture for the beam must be kept large enough.

### 2.3 Quality factor and stored energy

The quality factor  $Q$  is defined by:

$$Q = \frac{\omega W_s}{P_d}$$

where  $W_s$  is the stored energy. Clearly  $Q$  remains the same if the structure is used either in the standing-wave mode or the travelling-wave mode. It is also common to use the stored energy per unit length of the structure  $w_s = dW_s/dz$ .

Then

$$Q = -\frac{\omega w_s}{dP/dz}$$

Another quantity of interest is the ratio  $r/Q$ :

$$\frac{r}{Q} = \frac{E_z^2}{\omega w_s}$$

quantity which only depends on the cavity geometry at a given frequency, and which can be measured directly by a perturbation method. The other quantities depend on other factors like the wall material, the quality of brazing etc. ...  $Q$  varies like  $\omega^{-1/2}$ , hence  $r/Q$  varies like  $\omega$ .

Exercise Fields, quality factor Q and ratio r/Q for a pill-box cavity

Note that pill-box cavities are very representative of single-cell accelerating structures in most cases.

The field components for TM<sub>n,p,q</sub> modes in cylindrical cavities are given by:

$$\begin{aligned}
E_z &= k_2^2 \cos k_1 z \ J_n(k_2 r) \cos n\theta \\
E_r &= -k_1 k_2 \sin k_1 z \ J_n'(k_2 r) \cos n\theta \\
E_\theta &= \frac{n k_1}{r} \sin k_1 z \ J_n(k_2 r) \sin n\theta \\
H_z &= 0 \\
H_r &= -\frac{j}{Z_o} \frac{nk}{r} J_n(k_2 r) \sin n\theta \\
H_\theta &= -\frac{j}{Z_o} k k_2 J_n'(k_2 r) \cos n\theta \\
Z_o &= (\mu_o / \epsilon_o)^{1/2}
\end{aligned}$$

satisfying the boundary conditions:

$$\begin{aligned}
E_r = E_\theta &= 0 \quad \text{for } z = 0 \quad \text{and} \quad z = \ell \\
E_z &= E_\theta = 0 \quad \text{for } r = a
\end{aligned}$$

with

$$\begin{aligned}
k_1 &= \frac{q\pi}{\ell} \quad J_n(k_2 a) = 0 \quad k_2 = \frac{v_{np}}{a} \\
k^2 &= \frac{4\pi^2}{\lambda^2} = \left(\frac{q\pi}{\ell}\right)^2 + \left(\frac{v_{np}}{a}\right)^2
\end{aligned}$$

where  $v_{np}$  is the  $p^{\text{th}}$  root of  $J_n(x) = 0$  and  $\lambda$  the free-space wavelength.

The most simple mode in a cylindrical cavity is the mode  $\text{TM}_{010}$ . This is the fundamental mode which however requires  $l/a < 2$ . This mode has only two components (Fig. 15):

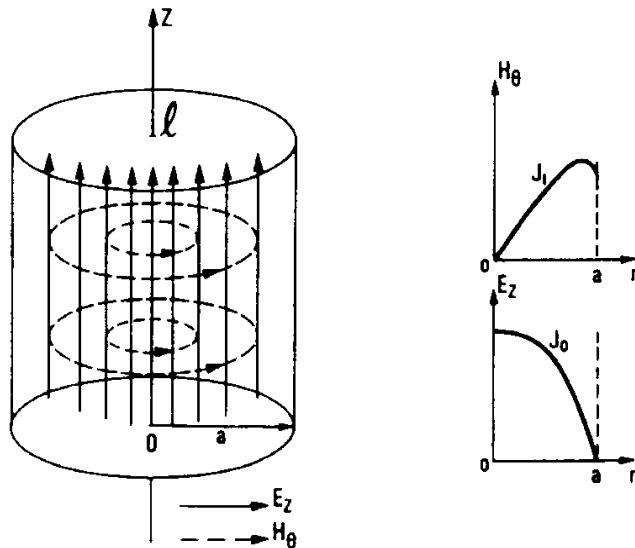


Fig. 15  $\text{TM}_{010}$  mode in a pill-box cavity

$$E_z = J_0(kr)$$

$$H_\theta = -\frac{j}{Z_0} J_1(kr)$$

$$(J'_0 = -J_1)$$

The resonant frequency is given by  $\nu_{np} = 2.4$  and  $\lambda = 2\pi a/2.4 = 2.62 a$ .

For  $\lambda = 10$  cm one gets  $a = 3.8$  cm.

In a resonant RLC circuit,  $Q$  is expressed as follows:

$$Q = 2\pi f \frac{\frac{1}{2}LI^2}{\frac{1}{2}RI^2} = \frac{L\omega_0}{R} \quad \text{with} \quad \omega_0 = \frac{1}{\sqrt{LC}} .$$

So, one can write for the definition of  $Q$

$$Q = 2\pi \frac{\text{Stored energy}}{\text{Energy lost during one period}}$$

which can now be extended to a resonant cavity.

The stored energy in the cavity volume is given by:

$$W_s = \frac{\mu}{2} \int_V |H|^2 dV = \frac{\epsilon}{2} \int_V |E|^2 dV .$$

For the power losses in the walls, one notices that the magnetic field induces in the wall a current  $\vec{i} = \vec{n} \times \vec{H}$  or  $i = H$ . Then the losses are given by:

$$P_d = \frac{1}{2} \int_S R_w H^2 dS$$

where  $R_w$  is the surface resistance for a layer of unit area and width  $\delta$  (skin depth):

$$R_w = \frac{1}{\sigma \delta} \quad \text{with} \quad \delta = \frac{1}{\sqrt{\pi \mu \sigma f}}$$

and where  $\sigma$  is the material conductivity and  $f$  the RF frequency. So:

$$dP_d = \frac{\pi \mu \delta}{2} |H^2| f dS .$$

The energy lost during one period is:

$$dW_d = \frac{1}{f} dP_d = \frac{\pi \mu \delta}{2} |H^2| dS$$

and for the total wall surface:

$$W_d = \frac{\pi\mu\delta}{2} \int_S |H|^2 dS .$$

Hence:

$$Q = \frac{2}{\delta} \frac{\int_V |H|^2 dV}{\int_S |H|^2 dS} = \frac{2}{\delta} \frac{KV}{S}$$

where  $K$  is the form factor of the given geometry.

Considering again the TM<sub>010</sub> mode in a pill-box cavity one gets:

$$\begin{aligned} \int_V H_\theta^2 dV &= \ell \int_0^a J_1^2(k_2 r) 2\pi r dr \\ \int_S H_\theta^2 dS &= 2 \int_0^a J_1^2(k_2 r) 2\pi r dr + 2\pi a \ell J_1^2(k_2 a) \end{aligned}$$

so

$$\frac{1}{Q} = \frac{\delta}{\ell} \frac{\int_0^a J_1^2(k_2 r) r dr + \frac{a\ell}{2} J_1^2(k_2 a)}{\int_0^a J_1^2(k_2 r) r dr} .$$

From the relation:

$$\int_0^a J_1^2(k_2 r) r dr = \frac{a^2}{2} J_1^2(k_2 a)$$

one gets

$$Q = \frac{\ell}{\delta} \frac{a}{a + \ell} \propto \omega^{-1/2}$$

and for example:

$$\delta = 10^{-6} \text{ m} \quad a = 3.8 \times 10^{-2} \text{ m} \quad \ell = 5 \times 10^{-2} \text{ m}$$

gives  $Q = 21590$ .

In addition one can also get the quantity  $r/Q$  ( $r$  being the uncorrected shunt impedance)

$$\frac{r}{Q} = \frac{V^2}{\omega W_s \ell} = 2.58 \text{ } \mu\text{f} \propto \omega$$

hence

$$r \propto \omega^{1/2} .$$

## 2.4 Filling time

From the definition of  $Q$  one has for a resonant cavity:

$$P_d = \frac{\omega}{Q} W_s .$$

If the cavity has been initially filled, the rate at which the stored energy decreases is related to the power dissipated in the walls:

$$\frac{dW_s}{dt} = -\frac{\omega}{Q} W_s .$$

Hence the time it takes for the electric field to decay to  $1/e$  of its initial value is:

$$t_f = \frac{2Q}{\omega}$$

which is the filling time of the cavity. In the case of a travelling-wave structure the definition of the filling time is different

$$t_f = \frac{L}{v_e}$$

where  $L$  is the length of the structure and  $v_e$  the velocity at which the energy propagates. In a travelling-wave structure the stored energy exists but never adds up because it is dissipated in a terminating load and does not reflect

## 2.5 Phase velocity and group velocity

These two concepts are of high importance in the case of particle acceleration by means of travelling guided waves. As mentioned before such methods are mostly used for particles whose velocity is either close or equal to the light velocity  $c$ .

Let's first assume a cylindrical waveguide, and search for the simplest TM (or E) mode which can propagate. Such a mode, with an axial electric field component  $E_z$ , is the TM<sub>01</sub> mode which also has two transverse components  $E_r$  and  $H_\theta$ :

$$\begin{aligned} E_z &= E_0 J_0(k_c r) e^{-j\beta z} \\ E_r &= j \frac{\beta}{k_c} E_0 J_1(k_c r) e^{-j\beta z} \\ H_\theta &= \frac{1}{Z_0} j \frac{k}{k_c} E_0 J_1(k_c r) e^{-j\beta z} \\ Z_0 &= \sqrt{\frac{\mu_0}{\epsilon_0}} = 377 \text{ ohms} \end{aligned}$$

where  $\beta$  is the propagation factor of the wave travelling in the  $+z$  direction, satisfying the relation:

$$\beta^2 = k^2 - k_c^2$$

with:

$$k = \frac{2\pi}{\lambda} = \frac{\omega}{c}$$

$$J_0(k_c a) = 0 \quad k_c a = 2.4$$

$$k_c = \frac{2\pi}{\lambda_c} = \frac{\omega_c}{c}$$

and where  $a$  is the inner radius of the cylindrical waveguide,  $\omega$  the excitation frequency,  $\omega_c$  the cut off frequency.

The wavelength  $\lambda_g$  of the propagating wave is such that:

$$\beta = \frac{2\pi}{\lambda_g} = \frac{\omega}{v_p}$$

where  $v_p$  is the velocity of the wave or phase velocity. In order for the wave to propagate  $\lambda_g$  must be real and positive which means:

$$\lambda < \lambda_c$$

or

$$v_p > c$$

In order to lower the phase velocity the waveguide is loaded by disks, equally spaced if the particle is ultra-relativistic ( $v \sim c$ ). The disks act like capacitive loads and reduce the speed of propagation as in loaded transmission lines.

It is usual to draw the Brillouin diagram for the type of propagating wave under consideration. This diagram relates the frequency to the propagation factor (Fig. 16).

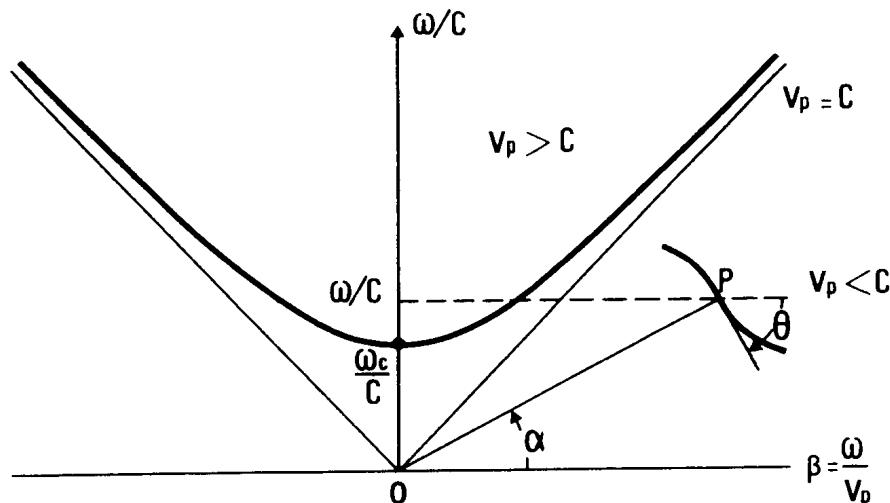


Fig. 16 Brillouin diagram

The straight line  $v_p = c$  separates the two domains corresponding respectively to slow and fast waves. For the latter, as obtained in a normal guide, the relation

$$\frac{\omega^2}{v_p^2} = \frac{\omega^2}{c^2} - \frac{\omega_c^2}{c^2}$$

gives a hyperbola for a given  $\omega_c$ .

For a slow wave it will exit an operating point P in the diagram and the corresponding phase velocity is given by  $\tan\alpha = v_p/c$ . If  $\omega$  varies, P moves on a certain curve; the slope of this curve at point P is:

$$\tan\theta = \frac{d(\omega/c)}{d(\omega/v_p)} = \frac{1}{c} \frac{d\omega}{d\beta} = \frac{1}{c} v_g$$

where  $v_g = (d\beta/d\omega)^{-1}$  is called the group velocity and happens to be equal to the velocity of the energy flow in the waveguide:

$$v_g = v_e$$

Exercise: Calculation of the energy flow velocity

The average power which flows through a transverse cross-section of a waveguide is given by the integral of the Poynting vector:

$$P = \frac{1}{2} \operatorname{Re} \int_S (E_T \times H_T) dS$$

where only the transverse components of the field have to be considered. For a TM mode the relation between  $E_T$  and  $H_T$  is:

$$\begin{aligned} \frac{E_T}{H_T} &= Z_0 \frac{\lambda}{\lambda_g} \\ P &= \frac{1}{2} \int_S \frac{1}{Z_0} \frac{k}{\beta} |E_T|^2 dS \end{aligned}$$

The energy stored in the magnetic field (purely transverse component) per unit length is:

$$w_{sm} = \frac{\mu}{4} \int_S |H_T|^2 dS = \frac{\mu}{4} \frac{1}{Z_0^2} \frac{k^2}{\beta^2} \int_S |E_T|^2 dS$$

The energy stored in the electric field per unit length is equal to that of the magnetic field. Hence the total stored energy per unit length is:

$$w_s = w_{se} + w_{sm} = 2w_{sm}$$

The velocity of the energy flow is then given by:

$$v_e = \frac{P}{w_s} = \frac{\frac{1}{2} Z_0 \frac{k}{\beta}}{\mu \frac{1}{2} \frac{k^2}{Z_0^2} \beta^2} = \frac{1}{\mu} Z_0 \frac{\beta}{k} = \frac{1}{\mu} \sqrt{\frac{\mu_o}{\epsilon_o}} \frac{\beta}{k}$$

$$v_e = \frac{\beta}{k} c$$

Since:

$$v_g = \left( \frac{d\beta}{d\omega} \right)^{-1} = c \frac{dk}{d\beta} = \left[ \frac{d(\omega^2 / c^2 - k_c^2)^{1/2}}{d\omega} \right]^{-1}$$

one gets:

$$v_g = \frac{\beta c^2}{\omega} = \frac{\beta c}{k} = v_e$$

## 2.6 Space harmonics in loaded waveguides

In an infinite periodic structure (Fig. 17) the wave equation must satisfy the periodic boundary condition imposed by the disks. This is obtained by choosing a solution of the form:

$$E(r, \theta, z) = e^{-\gamma z} E_1(r, \theta, z)$$

$$H(r, \theta, z) = e^{-\gamma z} H_1(r, \theta, z)$$

where  $E_1$  and  $H_1$  are periodic functions:  $E_1(r, \theta, z + d) = E_1(r, \theta, z)$ . Considering two similar terminal planes in two consecutive cells, that means the fields will repeat except for the multiplication factor  $e^{-\gamma d}$  which can be related to the propagation time from one cell to the next.

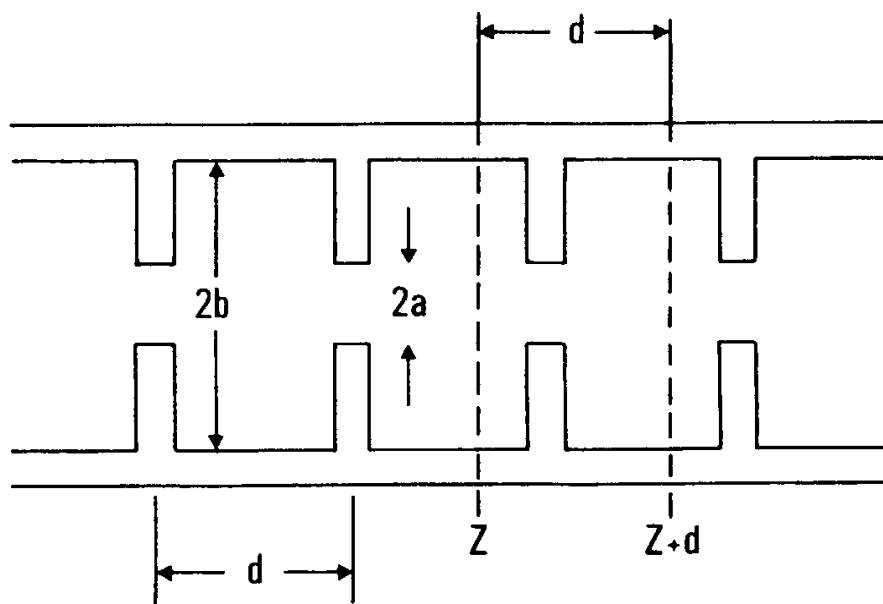


Fig. 17 Periodic loaded structure

The possibility of expressing the field in the above form is often referred to as Floquet's theorem (the original Floquet's theorem dealt with differential equations with periodic coefficients, the case of periodic boundary conditions is an extension of that work).

Any periodic function such as  $E_1(r, \theta, z)$  can be expanded into an infinite Fourier series; thus

$$E_1(r, \theta, z) = \sum_{n=-\infty}^{+\infty} E_{1n}(r, \theta) e^{-j2n\pi z/d}$$

Considering a lossless structure, the propagation requires  $\gamma$  to be imaginary:

$$\gamma = j\beta_o$$

Hence the field becomes:

$$E(r, \theta, z) = \sum_{n=-\infty}^{+\infty} E_{1n}(r, \theta) e^{-j\beta_n z}$$

with

$$\beta_n = \beta_o + 2n\pi / d$$

$\beta_o$  is the propagation factor of the fundamental space harmonic.

In addition the field has to satisfy the usual transverse boundary conditions of cylindrical waveguides. Thus, if one concentrates only on the lowest TM type mode the field components in a periodic disk-loaded structure are:

$$\begin{aligned} E_z &= \sum_n E_{on} J_0(k_{c,n} r) e^{-j\beta_n z} \\ E_r &= j \sum_n \frac{\beta_n}{k_{c,n}} E_{on} J_1(k_{c,n} r) e^{-j\beta_n z} \\ H_\theta &= \frac{j}{Z_o} \sum_n \frac{k}{k_{c,n}} E_{on} J_1(k_{c,n} r) e^{-j\beta_n z} \end{aligned}$$

with the more general relation

$$\beta_n^2 = k^2 - k_{c,n}^2$$

Notice that all the space harmonics exist at a given frequency  $\omega$ . Once  $\beta_o$  is known all  $\beta_n$ 's are known. Moreover each space harmonic has a different phase velocity given by:

$$v_{pn} = \frac{\omega}{\beta_o + \frac{2\pi n}{d}}$$

The group velocity of the  $n^{\text{th}}$  harmonic is:

$$v_{gn} = \frac{d\omega}{d\beta_n} = \left( \frac{d\beta_n}{d\omega} \right)^{-1} = \left( \frac{d\beta}{d\omega} \right)^{-1} = v_g$$

It is the same for all harmonics.

The Brillouin diagram for a loaded structure (or slow wave structure) is represented on Fig. 18. At a given frequency there is an infinite number of points P corresponding to the propagation factors  $\beta_n$ . If the frequency changes, the points move on curves which have all the same slope, corresponding to the group velocity. Hence one gets pieces of curve which can all be deduced by a simple translation of  $2\pi/d$ . Since these curves must join, obviously they must have zero slopes somewhere leading to the shape of Fig. 18. It happens that the first part of the curve has a zero slope, hence zero group velocity, at  $\beta d = 0$  and  $\pi$  which give the lower  $\omega_0/c$  and upper  $\omega_\pi/c$  frequencies of a pass band which remains true for the higher space harmonics. The condition of propagation hence corresponds to:

$$0 \leq \cos \beta d \leq 1$$

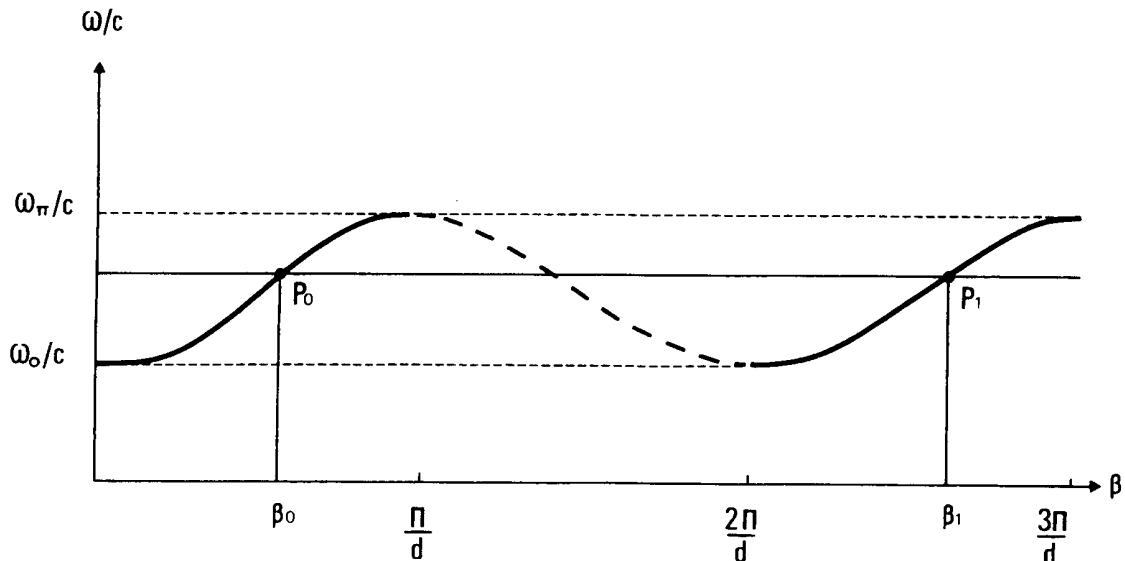


Fig. 18 Brillouin diagram for a slow wave structure

At  $\beta d = \pi$  the phases in two successive cells are opposite ( $\pi$  mode) and one gets a standing wave pattern in the combination of all the space harmonics.

The calculation of the real fields in loaded structures which would take careful account of all boundary conditions is tedious. Generally a reasonably accurate description of the dispersion curve, relating  $\beta$  to  $k$ , is obtained by an equivalent transmission line analysis or a coupled resonators chain analysis.

In order to accelerate electrons which already have the velocity of light  $c$  the operating point in the Brillouin diagram must correspond to the intersection of the  $k, \beta$  curve with the  $45^\circ$  line (Fig. 19).

In a travelling wave structure the  $\pi$  mode is avoided according to the previous remarks. Either  $\pi/2$  or  $2\pi/3$  modes are used. For the former a phase shift of  $2\pi$  is obtained over four cells while for the latter it needs three cells.

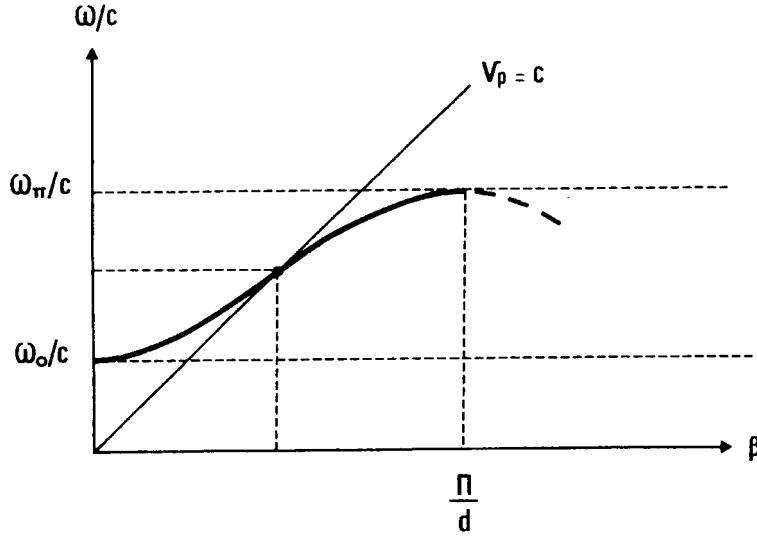


Fig. 19 Operating point for  $v_p = c$

For an acceleration to take place one chooses the fundamental space harmonic such that:

$$\begin{aligned}\beta_o &= \frac{\omega}{c} \\ \beta_o d &= \frac{\pi}{2} \quad \text{or} \quad \frac{2\pi}{3}\end{aligned}$$

The fundamental is the only one to give a net accelerating field since the higher space harmonics have no effect on the average. Hence, most of the particle dynamics in a travelling wave structure can be treated only using the fundamental space harmonic which can be calculated, for any structure having cylindrical symmetry, with the help of powerful computer codes like LALA or SUPERFISH. The radial dimension of the structure is determined to fit the operating frequency  $\omega$ .

### 3. ENERGY GAIN IN LINEAR ACCELERATING STRUCTURES

#### 3.1 Standing-wave structures

The energy gain in a standing-wave structure is straightforward when the shunt impedance corrected by the transit time factor is known. However, when the power source is matched to the resonant structure through a coupling loop, such that no power is reflected toward the source, then the loaded  $Q$  value becomes:

$$Q_L = \frac{Q}{1 + \beta}$$

where the coupling coefficient  $\beta$  is unity when the power given to the beam is negligible. The corresponding filling time now becomes:

$$t_f = \frac{2Q_L}{\omega} = \frac{2Q}{\omega(1 + \beta)}$$

For long power pulses there is no peculiar effect, but for short power pulses whose length is of the order of the filling time (acceleration of short beam pulses) there will be a transient effect during the filling of the cavity where reflected power cannot be avoided. Hence the minimum power required to attain a certain accelerating voltage will depend on the coupling coefficient and not only on the shunt impedance.

### 3.2 Travelling-wave structures

For travelling-wave structures the energy gain will depend on the RF characteristics of the cells in a more complicated way. It is usual to distinguish between constant-impedance structures and constant-gradient structures.

#### 3.2.1 Constant-impedance structure

In a constant-impedance structure, for instance of the disk-loaded type, all the cells are identical. Hence the group velocity  $v_g$ , which depends on the geometrical parameters of the cells, remains constant all along the structure. If  $L$  is the total length of the structure the filling time is:

$$t_f = \frac{L}{v_g}$$

For a disk-loaded structure such as the one drawn on Fig. 7, operating in the  $2\pi/3$  mode at 3 GHz, the group velocity, which is a strong function of the iris diameter  $2a$  is approximately given by:

$$v_g / c = \frac{(2a)^{3.23}}{891}$$

where  $2a$  is expressed in cm. For instance,  $v_g = 0.01 c$  for  $2a = 1.97$  cm.

As the waves propagate in such a structure, part of the input power is dissipated into the walls and the remaining power will lead to a smaller accelerating field. Hence the accelerating gradient decreases continuously along the structure. It was shown in the previous section that:

$$\begin{aligned} Q &= -\frac{\omega w_s}{dP/dz} \\ \frac{r}{Q} &= \frac{E^2}{\omega w_s} \\ w_s &= P/v_g \end{aligned}$$

So one can write:

$$\frac{dP}{dz} = -\frac{\omega P}{v_g Q}$$

leading to the exponential behaviour:

$$\begin{aligned} P &= P_0 e^{-(\omega/v_g Q)z} \\ E &= E_0 e^{-(\omega/2v_g Q)z} \end{aligned}$$

At the input of the structure the accelerating field  $E_o$  is related to the input power  $P_o$ :

$$E_o^2 = P_o \frac{\omega}{v_g} \frac{r}{Q}$$

Integrating the field along the structure leads to the energy gain:

$$V_{\text{total}} = \int_0^L E(z) dz = 2 E_o v_g \frac{Q}{\omega} \left[ 1 - e^{-(\omega/2v_g Q)L} \right]$$

It is common to use the attenuation factor  $\tau$  of the structure as follows:

$$e^{-2\tau} = P(z=L) / P(z=o)$$

Then

$$\tau = \frac{1}{2} \frac{\omega L}{Q v_g} = \frac{1}{2} \frac{\omega}{Q} t_f$$

and the expression for the energy gain becomes:

$$V_{\text{total}} = (P_o r L)^{1/2} \left[ (2\tau)^{1/2} (1 - e^{-\tau}) / \tau \right]$$

*Exercise:* Consider a disk-loaded structure, 1-meter long, operating at 3 GHz in the  $2\pi/3$  mode and assume an iris diameter of 1.97 cm which gives  $v_g/c = 0.01$ . For such a structure made of copper the shunt impedance per meter is roughly:

$$r_{[\text{M}\Omega/\text{m}]} = 86 - 3.6(2a)^2_{[\text{cm}]} = 72 \text{ M}\Omega / \text{m}$$

while  $Q$  is practically independant of  $(2a)$  and equal to 15000.

For this example the performances of the structure are:

$$t_f = 0.33 \text{ } \mu\text{s}$$

$$\tau = 0.21 \text{ neper}$$

$$V_{[\text{MeV}]} = 5 \sqrt{P_{o[\text{MW}]}}$$

### 3.2.2 Constant-gradient structure

In order to compensate for the variation of the accelerating field along the structure, due to power dissipation, it is possible to lower the group velocity from cell to cell, by changing the geometry for the same operating frequency. In fact this is obtained by reducing the iris aperture and by reducing the diameter of the cell at the same time. Such a scheme will make a better use of the available power.

In a perfect constant-gradient structure  $E = cte$ , so one must have:

$$\frac{dP}{dz} = cte$$

assuming the shunt impedance is not too much affected by the change in the iris aperture.

Then one can write:

$$P = P_o - (P_o - P_L) \frac{z}{L}$$

where  $L$  is the length of the structure,  $P_o$  the input power and  $P_L$  the output power.

By analogy with the constant impedance case it is usual to define the attenuation factor  $\tau$  such that:

$$P_L / P_o = e^{-2\tau}$$

Then one has:

$$\begin{aligned} P &= P_o \left[ 1 - (1 - e^{-2\tau}) \frac{z}{L} \right] \\ \frac{dP}{dz} &= -P_o (1 - e^{-2\tau}) / L = -\frac{\omega P}{v_g Q} \end{aligned}$$

leading to a linear variation for the group velocity:

$$v_g(z) = \frac{\omega L}{Q} \frac{\left[ 1 - (1 - e^{-2\tau}) \frac{z}{L} \right]}{(1 - e^{-2\tau})}$$

The values of the group velocity at both extremities are:

$$\begin{aligned} v_g(0) &= \frac{\omega L}{Q} \frac{1}{1 - e^{-2\tau}} \\ v_g(L) &= \frac{\omega L}{Q} \frac{e^{-2\tau}}{1 - e^{-2\tau}} \end{aligned}$$

The attenuation factor is well defined when the output group velocity is known.

$$e^{2\tau} = 1 + \frac{\omega L}{Q v_g(L)}$$

In practice the iris diameter at the end of the structure is made as small as possible compatible with the dimensions of the accelerated beam. Setting the output group velocity and the length of the structure give the input group velocity.

The filling time in the present case is:

$$\begin{aligned} t_f &= \int_0^L \frac{dz}{v_g(z)} = \frac{Q}{\omega} L n \left[ 1 + \frac{\omega L}{Q v_g(L)} \right] \\ t_f &= 2\tau \frac{Q}{\omega} \end{aligned}$$

Integrating the accelerating field gives the energy gain:

$$V_{\text{total}} = \left[ P_o r L (1 - e^{-2\tau}) \right]^{1/2}$$

with

$$E = \left[ \frac{P_o r}{L} (1 - e^{-2\tau}) \right]^{1/2}$$

*Exercise:* Let's take the previous set of parameters and keep the shunt impedance constant. One gets:

$$\begin{aligned} v_g(L) / c &= 0.01 \\ \tau &= 0.175 \text{ neper} \\ t_f &= 0.28 \text{ } \mu\text{s} \\ v_g(o) / c &= 0.014 \\ V_{[\text{MeV}]} &= 4.6 \sqrt{P_{o[\text{MW}]}} \end{aligned}$$

The constant-gradient case appears to be slightly less efficient. However in the constant-impedance case the maximum field which takes place at the input is higher for equal energy gains. Since there is some worry about field breakdown on the walls, the constant-gradient structure finally appears more interesting although it is more difficult to build. An alternative is to build quasi-constant-gradient structures made of constant impedance landings, with transition cells between the landings. In that case the number of different cells is reduced.

For travelling-wave linacs the length of the power pulse must be at least equal to the filling time in order to accelerate very short bunches. The particles travel fast through the structure as compared to the group velocity so they must enter when the structure is completely filled. For longer beam pulses the power pulse must follow in length.

Pulsed klystrons are available at the level of 50 MW with a pulse length  $\leq 5 \text{ } \mu\text{s}$ . For short bunches ( $\leq 10 \text{ ns}$ ) compression schemes are used which give shorter power pulses ( $\leq 1 \text{ } \mu\text{s}$ ), compatible with the structure filling time, with a higher peak power which can be as much as 4 times the direct klystron peak power. This is either used to double the energy of existing linacs (for instance SLAC at Stanford) or to reduce the total number of power sources for a given nominal energy (for instance the LEP Injector Linac).

#### 4 . PARTICLE DYNAMICS IN LINEAR ACCELERATORS

Up to now a synchronism condition has been defined as a necessary condition for the particles to be accelerated in a linear structure. However this simple approach to the problem is not sufficient as it only describes the behaviour of selected particles which enter the accelerator at the right time with the right velocity.

In fact a bunch of particles, as it is produced by the gun, has a spread in velocities as well as a finite transverse dimension (beam emittance). Depending on their initial conditions the particles will undergo different stories during the acceleration. It is usual to differentiate the transverse motion of the particles from their longitudinal motion.

#### 4.1 Longitudinal motion: phase stability

If particles enter continuously in an accelerating structure obviously a fraction of these particles will see the axial field  $E_z$  at the wrong time (or wrong phase) due to the sinusoidal time variation. This already gives a feeling of the bunching phenomenon that will occur in the two types of accelerators (discrete acceleration through gaps or continuous acceleration with travelling waves).

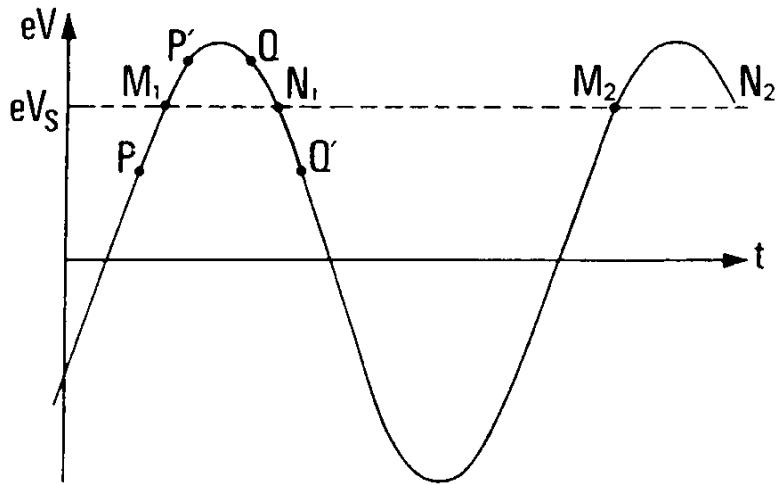


Fig. 20 Particle phases relative to the RF field

Two particles which arrive at different times of the accelerating half period (see Fig. 20) can be either subject to equal energy gains ( $M_1, N_1$ ) or different energy gains ( $P, P'$ ). Considering for instance an Alvarez structure in which the synchronism condition is obtained for a given energy gain  $eV_s$ , then particles  $M_1, N_1, M_2$  etc. will always see the same phase of the accelerating field.

A particle  $P$  which arrives in a gap in advance as compared to  $M_1$  will get less energy and its velocity will be smaller so that it will take more time to travel through the drift tube. In the next gap it will appear closer to particle  $M_1$ . The effect is true for particle  $P'$  which will get more energy and reduce its delay time as compared to  $M_1$ . Points  $M_1, M_2$  etc. are stable points for the acceleration since particles slightly away from them will experience forces that will reduce their deviation. On the contrary it can be seen that points  $N_1, N_2$  etc. are unstable points in the sense that particles slightly away from these points will shift even more in the next gaps.

In order to study the longitudinal motion one uses variables which give relative position, and energy, as compared to the synchronous particle:

$$\begin{aligned}\Delta t &= t - t_s \\ \varphi &= \phi - \phi_s = \omega(t - t_s) = \omega\Delta t \\ w &= W - W_s\end{aligned}$$

where  $\phi_s$  is the RF phase of the synchronous particle and  $W_s$  the energy of the same particle. The accelerating field can be simply described by:

$$E_z = E_0 \cos \left( \omega t - \frac{\omega z}{v_p} \right) = E_0 \cos \phi$$

When the phase velocity varies,  $\omega z/v_p$  must be replaced by  $\omega \int \frac{dz}{v_p}$  in the forthcoming expressions so that they will remain valid. The azimuthal position  $z$  is generally taken as the independant variable instead of  $t$ .

#### 4.1.1 Non-relativistic case – Adiabatic damping

The rate of energy gain for the synchronous particle is given by:

$$\frac{dW_s}{dz} = \frac{d}{dz} \left( \frac{1}{2} mv_s^2 \right) = \frac{d}{dt} (mv_s) = eE_o \cos \phi_s$$

where  $E_o$  would take into account the transit time in the case of standing-wave drift-tube linacs.

For other particles the energy gain can be expressed in reduced variables:

$$\frac{dw}{dz} = eE_o [\cos(\phi_s + \varphi) - \cos \phi_s]$$

In addition one has:

$$\frac{d\varphi}{dz} = \omega \left( \frac{dt}{dz} - \frac{dt_s}{dz} \right) = \omega \left( \frac{1}{v} - \frac{1}{v_s} \right)$$

which turns out to be:

$$\frac{d\varphi}{dz} = -\frac{\omega}{mv_s^3} w$$

having assumed that:

$$w = W - W_s = \frac{1}{2} m (v^2 - v_s^2) \cong mv_s (v - v_s)$$

Considering small deviations from the synchronous particle one gets the following set of first-order linear equations:

$$\begin{aligned} \frac{d\varphi}{dz} &= -\frac{\omega}{mv_s^3} w \\ \frac{dw}{dz} &= -eE_o \sin \phi_s \varphi \end{aligned}$$

which respresent a phase harmonic oscillation:

$$\frac{d^2\varphi}{dz^2} + \Omega_s^2 \varphi = 0$$

with angular frequency relative to the independant variable  $z$ :

$$\Omega_s^2 = -\frac{eE_o \omega \sin \phi_s}{mv_s^3}$$

showing that  $\sin \phi_s$  has to be negative for stable solutions. Notice that this is true since  $v_s$  is a constant parameter. But the approach is still valid if  $v_s$  varies slowly. The wavelength of the small amplitude phase oscillation is given by:

$$\lambda_s = 2\pi/\Omega_s$$

and increases rapidly along the accelerator.

Considering larger amplitudes one should write:

$$\frac{d^2\varphi}{dz^2} = -\frac{\omega}{mv_s^3} eE_0 [\cos(\phi_s + \varphi) - \cos \phi_s] = F$$

The restoring force  $F$  can be derived from an effective potential energy function  $U$

$$U = - \int F \, d\varphi$$

which is drawn on Fig. 21. An analysis of the non-linear motion in the phase space  $w, \varphi$  shows that the oscillation is bounded at some energy  $w_{\max}$ . The corresponding curve is called the separatrix.

The motion can be derived from a Hamiltonian  $H$  satisfying the canonical equations:

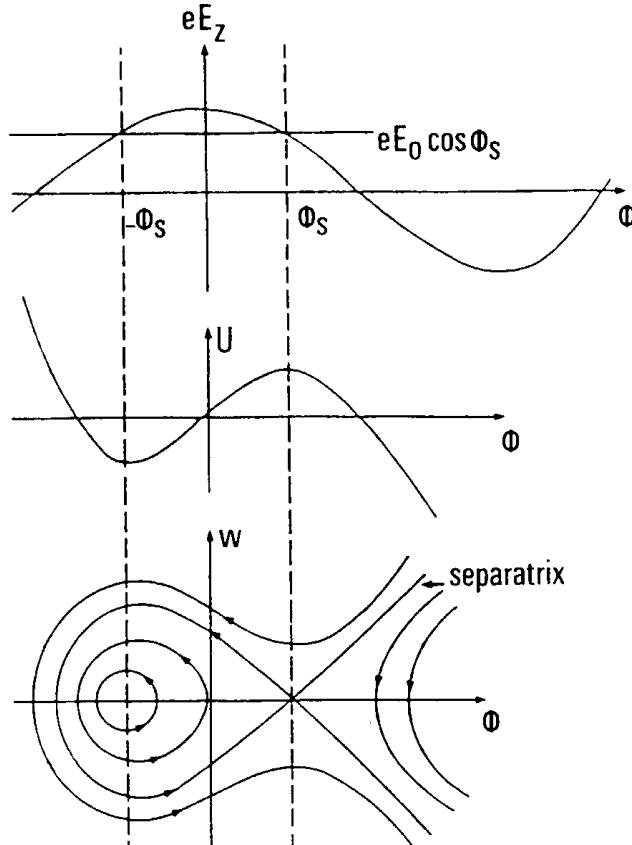


Fig. 21 Phase stability graphs

$$\frac{dw}{dz} = -\frac{\partial H}{\partial \varphi} \quad \frac{d\varphi}{dz} = \frac{\partial H}{\partial w}$$

One gets:

$$H = -\frac{\omega}{2mv_s^3} w^2 - eE_0 (\sin(\varphi + \phi_s) - \varphi \cos \phi_s)$$

which, for small amplitudes, reduces to:

$$H \approx -\frac{\omega}{2mv_s^3} w^2 - \frac{1}{2} eE_0 \sin \phi_s \varphi^2$$

A particle with some initial conditions will perform an ellipse in the phase space. Its maximum energy  $w_{\max}$  is obtained when  $\varphi = 0$  and correspondingly its maximum phase excursion  $\varphi_{\max}$  is obtained when  $w = 0$ . One has the relation:

$$\frac{w_{\max}}{\varphi_{\max}} = \left[ \frac{eE_0 \sin \phi_s mv_s^3}{\omega} \right]^{1/2}$$

Liouville's theorem stipulates that for a conservative motion the area of the phase space ellipse is an invariant

$$w_{\max} \varphi_{\max} = cte$$

This is normally true for a constant  $v_s$ , but can still be applied when considering adiabatic variation of this parameter (adiabatic theorem).

It follows from the previous relations that:

$$w_{\max} = cte \left( \frac{eE_0}{\omega} \sin \phi_s mv_s^3 \right)^{1/4}$$

$$\varphi_{\max} = cte \left( \frac{eE_0}{\omega} \sin \phi_s mv_s^3 \right)^{-1/4}$$

It appears that  $\varphi_{\max}$  will decrease during the acceleration and the bunch length will become shorter. On the contrary  $w_{\max}$  will increase, but the relative energy spread  $w/W_s$ , in which the user is interested, will decrease.

#### 4.1.2 Relativistic case – Electron capture

In the case of relativistic particles:

$$\frac{dt}{dz} = \frac{1}{v} = \frac{1}{c} \gamma (\gamma^2 - 1)^{-1/2}$$

with

$$\gamma = (1 - \beta^2)^{-1/2} \quad \beta = \frac{v}{c}$$

Then it happens that the previous formulae remain valid just replacing  $v_s$  by  $\gamma_s v_s$  and taking  $m$  as the rest mass.

The same phenomena occur for relativistic particles but the phase oscillation becomes very slow at high energies, so that in practice a particle can travel all along a linac in less than one oscillation period and some of the previous statements cannot have the same meaning any more. For instance if a short bunch of particles is captured in an electron linac and if all the particles have the light velocity they will keep that velocity and always arrive at the same RF phase in the gaps. In that case the phase spread of the bunch will remain constant. Moreover if this phase spread is very small, for instance around the peak of the RF, all the particles will get the same energy and the absolute energy spread will also remain constant while the relative energy spread will go down like  $\gamma^{-1}$ .

It is now interesting to consider the case of electron linacs, where the structure has a phase velocity equal to  $c$ , and to look at what happens when particles enter the structure having still a velocity smaller than  $c$ , as for instance produced by electron guns.

So let's consider an accelerator with a phase velocity equal to  $c$  and an axial electric field  $E_z$  which has a sinusoidal time variation with constant amplitude  $E_0$  along the trajectory. Then if  $E_z$  is the field seen by a particle one defines the phase angle  $\phi$  between the wave and the particle by means of the relation:

$$E_z = E_0 \sin \phi$$

If  $v$  is the particle velocity, then

$$d\ell = (c - v)dt$$

is the difference in path between the wave and the particle in the time  $dt$ . This path difference can also be expressed in term of the phase difference  $d\phi$ :

$$d\ell = \frac{\lambda_g}{2\pi} d\phi$$

where  $\lambda_g$  is the wavelength of the propagation in the waveguide. From the above equations one gets:

$$\frac{d\phi}{dt} = \frac{2\pi c}{\lambda_g} (1 - \beta)$$

The equation of motion for the particle, in the relativistic case, is simply:

$$\frac{d}{dt}(mv) = m_0 c \frac{d}{dt} \left[ \frac{\beta}{(1 - \beta^2)^{1/2}} \right] = eE_0 \sin \phi$$

and using a new variable  $\alpha$  such that:

$$\beta = \cos \alpha$$

it becomes

$$\frac{d\alpha}{dt} = -\frac{eE_0}{m_0 c} \sin \phi \sin^2 \alpha$$

with

$$\frac{d\phi}{dt} = \frac{d\phi}{d\alpha} \frac{d\alpha}{dt}$$

Hence:

$$-\sin \phi d\phi = \frac{2\pi m_o c^2}{\lambda_g e E_o} \frac{(1 - \cos \alpha)}{\sin^2 \alpha} d\alpha$$

Integrating from time  $t_1$  where  $v = c$ , to time  $t_0$  where  $\beta = \beta_o$ ,  $\alpha = \alpha_o$  and  $\phi = \phi_o$ , one gets:

$$\cos \phi_1 - \cos \phi_o = \frac{2\pi m_o c^2}{\lambda_g e E_o} \left[ \operatorname{tg} \frac{\alpha_1}{2} - \operatorname{tg} \frac{\alpha_o}{2} \right]$$

Knowing that:

$$\operatorname{tg} \frac{\alpha}{2} = \left( \frac{1 - \cos \alpha}{1 + \cos \alpha} \right)^{1/2} = \left( \frac{1 - \beta}{1 + \beta} \right)^{1/2}$$

one gets:

$$\cos \phi_o - \cos \phi_1 = \frac{2\pi m_o c^2}{\lambda_g e E_o} \left( \frac{1 - \beta_o}{1 + \beta_o} \right)^{1/2}$$

Since the left hand side cannot be greater than 2 one must have:

$$E_o \geq \frac{\pi m_o c^2}{\lambda_g e} \left( \frac{1 - \beta_o}{1 + \beta_o} \right)^{1/2}$$

This "capture condition" must be satisfied in order to capture the electrons injected at a velocity less than the phase velocity. For example, for  $\lambda_g = 10$  cm and an injection energy of 150 keV the condition gives  $E_o \geq 7.6$  MV/m which is technically possible.

In practice however, to improve the capture efficiency it is common to use a small bunching section, with a variable phase velocity, in which the energy is brought to a few MeV.

## 4.2 Transverse motion: defocusing

Looking at the electric field pattern in the gap between two drift tubes it is seen that there are radial components (Fig. 22), which are focusing at the gap entrance and defocusing at the end. In an electrostatic accelerator where the field is constant this gives a global focusing effect since the particle having more energy at the end of the gap makes the defocusing effect smaller.

In an RF accelerator the behaviour is different. From the phase stability requirement ( $\phi_s < 0$ ) it appears that the field increases with time during the passage of the particle. Hence the defocusing force becomes larger than the focusing one resulting in a transverse instability as the particle may strike the drift tubes.

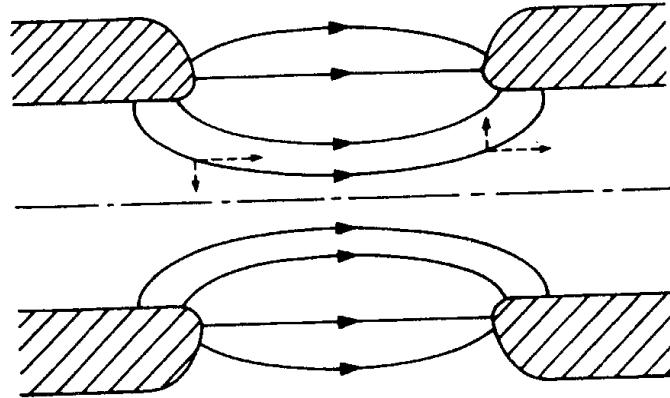


Fig. 22 Field pattern in the gap of a drift tube accelerator

It is possible to show the effect mathematically using approximate transverse field expressions only valid for small transverse deviations from the axis:

$$E_z = E_o \cos \left( \omega t - \omega \int \frac{dz}{v_p} \right)$$

$$E_r = -\frac{r\omega E_o}{2v_p} \sin \left( \omega t - \omega \int \frac{dz}{v_p} \right)$$

$$B_\theta = -\frac{r\omega E_o}{2c^2} \sin \left( \omega t - \omega \int \frac{dz}{v_p} \right)$$

The transverse force acting on the particle is given by the Newton-Lorentz equation:

$$\begin{aligned} \frac{d}{dt}(m\dot{r}) &= eE_r - evB_\theta \\ &= -\frac{er\omega E_o}{2v_p} \left( 1 - \frac{vv_p}{c^2} \right) \sin \left( \omega t - \omega \int \frac{dz}{v_p} \right) \end{aligned}$$

Considering the synchronous particle for which  $v = v_p$  and

$$\left( \omega t - \omega \int \frac{dz}{v_p} \right) = \phi_s$$

one gets:

$$\frac{d}{dt}(m\dot{r}) = -\frac{er\omega E_o}{2v_p} (1 - \beta^2) \sin \phi_s$$

For ultra-relativistic particles ( $v \approx c$ ) the transverse defocusing effect becomes negligible. In other words the transverse magnetic force compensates exactly the electric one. This is a well known behaviour in relativistic particle dynamics.

There are many other reasons for radial deflection of the particles than the normal transverse field components in a waveguide. For instance, in the input and output couplers of a travelling-wave structure there is an axial field asymmetry which induces transverse components and gives a transverse kick to the beam, even at  $v = c$ .

So finally in practice it is necessary to use external magnetic fields, such as those given by solenoids or quadrupoles, to ensure a stable transverse motion within the aperture of the linac.

### 4.3 Dynamics in a radio frequency quadrupole (RFQ)

Conventional proton (or heavy ion) linear structures, like the Alvarez structure, which use magnetic focusing, are only efficient in the range:

$$0.04 < \beta = \frac{v}{c} < 0.5$$

For  $\beta > 0.5$  it is better to use high- $\beta$  linac structures. For  $\beta < 0.04$  successful activity in designing low- $\beta$  structures is very recent (in the past, the solution consisted of applying a very high voltage on the gun to extract particles at energies that could match a drift-tube structure). The biggest success has been the invention of the radio frequency quadrupole (RFQ) which combines three functions: electric focusing, bunching and acceleration.

The RFQ is a four-vanes resonator with quadrupolar symmetry which provides a transverse electric gradient for transverse focusing (at low velocity, magnetic focusing is not efficient because of the  $v$  term which appears in the force equation). Modulated pole shapes (Fig. 23) lead to a longitudinal variation of the transverse field gradient giving a longitudinal electric component for acceleration and bunching. In a drift tube structure the transit time factor is worse at low  $\beta$ ; in the RF quadrupole many cells are made (since  $\beta$  is small) in an overall practical length which permit a continuous acceleration and perfect adiabatic conditions to produce a very good bunching efficiency ( $\sim 100\%$ ).

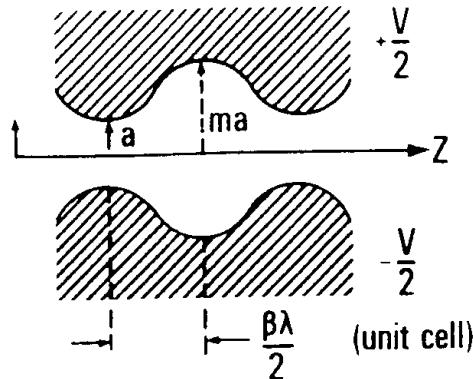


Fig. 23 Modulated pole shapes in an RFQ

The spatially continuous fields also cure the space charge effects that could freely develop in the drift tubes of conventional structures.

The lowest-order potential function, in cylindrical coordinates, which satisfies the quadrupole symmetry of an RFQ can be written as follows:

$$U = \frac{V}{2} \left[ X \left( \frac{r}{a} \right)^2 \cos 2\psi + A I_0(kr) \cos kz \right] \sin(\omega t + \phi)$$

where  $V$  is the difference potential between adjacent pole tips, and:

$$k = 2\pi/\beta\lambda$$

From this, the following electric field components are obtained:

$$\left. \begin{aligned} E_r &= -\frac{XV}{a^2} r \cos 2\psi - \frac{kAV}{2} I_1(kr) \cos kz \\ E_\psi &= \frac{XV}{a^2} r \sin 2\psi \\ E_z &= \frac{kAV}{2} I_0(kr) \sin kz \end{aligned} \right\} \sin(\omega t + \phi)$$

with

$$\begin{aligned} A &= (m^2 - 1) / (m^2 I_0(ka) + I_0(mka)) \\ X &= 1 - AI_0(ka) \end{aligned}$$

The quantity  $VA$  is the potential difference that exists on the axis between the beginning and the end of a unit cell. Then the space average longitudinal field is:

$$E_o = \frac{2AV}{\beta\lambda}$$

The energy gain for a particle with charge  $e$  and synchronous velocity  $\beta c$  traversing a unit cell is approximately:

$$\begin{aligned} \Delta W &= eE_o \ell T \cos\phi_s \\ \ell &= \frac{\beta\lambda}{2} \end{aligned}$$

and  $T = \pi/4$  is the transit time factor for a longitudinal field with space variation  $\sin kz$  and time variation  $\sin \omega t$ , knowing that when the particle travels one period  $\beta\lambda = 2\pi/k$ , the RF phase changes by  $2\pi$ .

Applying the equations of phase oscillations to the RFQ gives the angular frequency for small oscillations:

$$\Omega_s^2 = \frac{eAV\omega^2 \sin\phi_s}{4m_0c^2\beta^2}$$

the angular length of the separatrix,  $\phi_m$ :

$$\tan\phi_s = \frac{\sin\phi_m - \phi_m}{1 - \cos\phi_m}$$

and the spatial length of the separatrix:

$$z_m = \frac{\beta \lambda \phi_m}{2\pi}$$

Note that here  $\phi_m$  is the phase difference between the two extreme elongations of the separatrix.

To avoid space-charge phenomena it is interesting to keep the longitudinal density constant during bunching which means:

$$z_m = cte$$

condition which determines  $\phi_s$  and  $A$  as functions of  $\beta$ .

The next interesting aspect of the RFQ is the transverse focusing during bunching and acceleration. The magnitude of the electric quadrupole strength is  $XV/a^2$  which means that:

- for given  $a, m, \beta$  the strength is constant in a unit cell
- the same strength can be maintained in every unit by keeping  $XV/a^2$  constant.

The equation for the transverse oscillation is:

$$\frac{d^2x}{d\tau^2} = \left[ \frac{XV}{m_0 c^2} \left( \frac{\lambda}{a} \right)^2 \cos 2\pi\tau - \frac{\pi^2}{2} \frac{AV}{m_0 c^2 \beta^2} \sin \phi \right] x$$

where  $\tau$  is the reduced variable:

$$\tau = (\omega t + \phi) / 2\pi$$

The previous equation is of the Mathieu type:

$$\frac{d^2x}{d\tau^2} + [A + B \cos 2\pi\tau]x = 0$$

which means that solutions can be stable for some combinations of  $A$  and  $B$ . It can be shown that the oscillations will be stable if the following conditions are satisfied:

$$\begin{aligned} \frac{XV}{2\pi^2 m_0 c^2} \left( \frac{\lambda}{a} \right)^2 &< 1 - \frac{VA}{2m_0 c^2 \beta^2} \sin \phi \\ \frac{X^2 V}{8\pi^3 m_0 c^2} \left( \frac{\lambda}{a} \right)^4 &> -\frac{\pi A}{2\beta^2} \sin \phi \end{aligned}$$

An analysis of these inequalities shows that in practice the transverse oscillations are stable with any values of the synchronous phase (up to  $-90^\circ$ ) just as in accelerators with static lenses.

In practice an RFQ can bring proton (or ion) energies from a few 10 keV to a few MeV over a very reasonable overall length (1 to 2 meters).

## BIBLIOGRAPHY

- M.S. Livingston, J.P. Blewett, Particle Accelerators, McGraw-Hill Book Company, INC, N.Y. 1962
- E. Persico, E. Ferrari, S.E. Segré, Principles of Particle Accelerators, W.A. Benjamin, INC, N.Y. 1968
- P. Lapostolle, A. Septier (Editors), Linear Accelerators, North Holland Publishing Company, Amsterdam 1970
- P. Lapostolle, Introduction à l'étude des Accélérateurs Linéaires, CERN 66-20, 1966
- P.B. Wilson, High Energy Electron Linacs: Applications to Storage Ring RF Systems and Linear Colliders, SLAC PUB 2884. Also in AIP Conference Proceedings N°87, N.Y. 1982
- G.A. Loew, Elementary Principles of Linear Accelerators, SLAC PUB 3221. Also in AIP Conference Proceedings N°105, N.Y. 1983
- I.M. Kapchinskii and V.A. Teplyakov, Linear Ion Accelerator With Spatially Homogeneous Strong Focusing, Pribory i. Tekhnika Eksperimenta, 119, N°2 (March-April 1970)
- R.H. Stokes, K.R. Crandall, J.E. Stovall and D.A. Swenson, RF Quadrupole Beam Dynamics, IEE Transactions on Nuclear Science, Vol. NS-26, N°3 June 1979
- R.E. Collin, Foundations for Microwave Engineering, International-Student Edition, McGraw-Hill, INC. 1966

# LONGITUDINAL BEAM DYNAMICS IN CIRCULAR ACCELERATORS

J. Le Duff

Laboratoire de l'Accélérateur Linéaire  
Université Paris-Sud, 91405 Orsay, France

## 1. ACCELERATION BY TIME VARYING FIELDS

### 1.1 Time varying magnetic field: the betatron concept

The betatron accelerator is the only circular machine which uses a time varying magnetic field to accelerate the particle. It is typically an induction accelerator. Notice that induction linear accelerators are also used.

The betatron, as shown on Fig. 1, is a very simple machine which consists of a magnet fed by an alternating current at a frequency usually between 50 to 200 Hz. The magnet poles which surround the vacuum chamber where particles circulate are truncated cones. In this machine the magnetic field is used for guiding the particles on a circular trajectory as well as for acceleration.

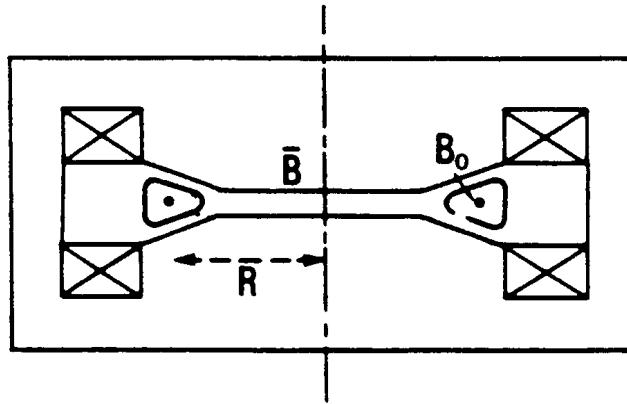


Fig. 1 The betatron scheme

The variable magnetic field produces an electric field component according to Maxwell's equations:

$$E = - \text{grad}V - \frac{\partial A}{\partial t}$$
$$B = \mu H = \text{curl } A$$

In the present case where there is no scalar potential, and according to the field symmetry, one has:

$$V = 0$$
$$A_r = A_z = 0 \quad A_\theta = A(r, z, t)$$
$$E_\theta = -\frac{\partial A_\theta}{\partial t}$$

showing that the electric field is tangent to the circular orbit defined by

$$B_o R = -p/e$$

where  $e$  is the algebraic particle charge.

Moreover, from the single component  $A_\theta$  one gets the magnetic field components:

$$\begin{aligned} B_r &= -\frac{\partial A}{\partial z} \\ B_\theta &= 0 \\ B_z &= \frac{1}{r} \frac{\partial}{\partial r} (rA) = A' + \frac{A}{r} \end{aligned}$$

The flux of  $B$  linking any circle  $r = cte$  is then:

$$\Phi = \int_0^r B_z 2\pi r dr = 2\pi \int_0^r \frac{1}{r} \frac{\partial}{\partial r} (rA) r dr = 2\pi r A .$$

Denoting by  $\bar{B}_z$  the average field value inside the circle, the flux will be also:

$$\Phi = \pi r^2 \bar{B}_z$$

which shows that:

$$A = \frac{1}{2} r \bar{B}_z .$$

Then one gets:

$$E_\theta = -\frac{\partial A}{\partial t} = -\frac{1}{2} \frac{\partial}{\partial t} (r \bar{B}_z) = -\frac{r}{2} \frac{\partial}{\partial t} \bar{B}_z .$$

Putting  $r = R$  the voltage over one turn is:

$$2\pi R E_\theta = -\pi R^2 \frac{d\bar{B}_z}{dt} = -\frac{d\Phi}{dt}$$

which is a well known law in electronics that a time variation of the flux induces a voltage. The induction accelerator is often considered as a transformer in which the primary current is the alternating current and the secondary current is the circulating electron beam.

The acceleration is given by the Newton-Lorentz equation:

$$\begin{aligned} \frac{dp}{dt} &= e E_\theta \\ \frac{dp}{dt} &= -\frac{e}{2\pi R} \frac{d\Phi}{dt} = -\frac{1}{2} e R \frac{d\bar{B}_z}{dt} \end{aligned}$$

If one wants to keep the particle on the same trajectory the following relation must also be fulfilled:

$$\frac{dp}{dt} = -e R \frac{dB_0}{dt}$$

and hence

$$B_0 = \frac{1}{2} \bar{B}_z + cte$$

which is the betatron relation.

## 1.2 Time varying electric field: the cyclotron case

The cyclotron is an example of a circular accelerator using a radio frequency accelerating field. A magnet (Fig. 2a) produces a constant  $B$  field in the useful area for the particle trajectories. The box in which the particles circulate is divided into two pieces, separated by a gap (Fig. 2b) fed with an RF signal which produces an alternating electric field.

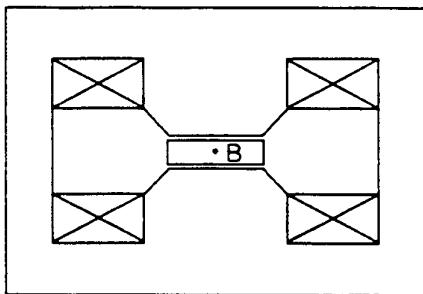


Fig. 2a Cyclotron magnet

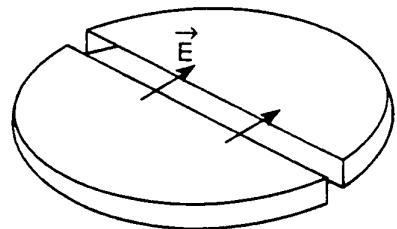


Fig. 2b Shape of electrodes

Each time the particle crosses the gap it experiences an accelerating force provided the phase of the RF has the right polarity. After each crossing the particle, with velocity  $v$ , follows a circular trajectory with radius  $r$ , due to the vertical magnetic field perpendicular to the particle velocity:

$$r = \frac{mv}{eB}$$

Having described half a circle the particle comes back through the gap and is accelerated once more provided the field polarity has reversed. Since the magnetic field is constant the particle trajectory will spiral as shown in Fig. 3.

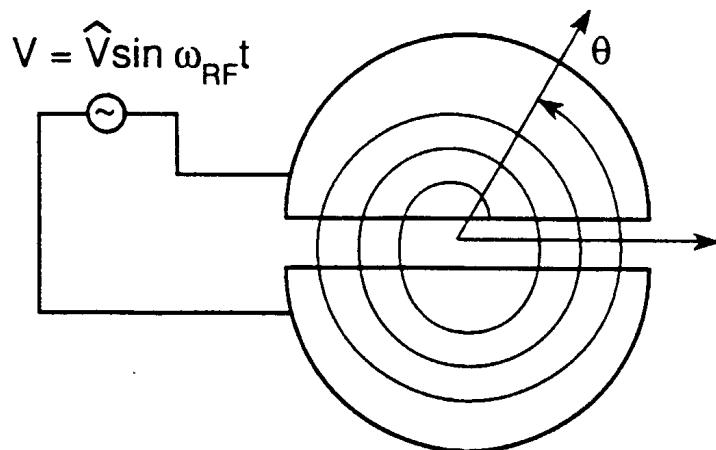


Fig. 3 Cyclotron orbit

A synchronism condition for the accelerator is obtained by proper choice of the RF frequency:

$$\omega_{RF} = \omega_r = \frac{eB}{m}$$

where  $\omega_r$  is the angular revolution frequency and  $m$  the mass. However, since standard cyclotrons have a fixed RF frequency, this relation can only be satisfied for constant mass, which implies non-relativistic particles:

$$m = m_0 .$$

This would be the case for instance, for heavy particles (protons and ions) for which the kinetic energy remains small compared to the rest energy.

Within the previous assumptions the cyclotron would remain a rather low energy machine. So let us consider the case of higher energies where particles become relativistic. At each transit in the gap the energy gain is:

$$\Delta E = e\hat{V} \sin \phi$$

where  $\phi$  is the phase of the particle with respect to the RF phase:

$$\phi = \omega_{RF}t - \theta$$

$\theta$  being the azimuthal angle of the trajectory. Differentiating with respect to time gives:

$$\dot{\phi} = \omega_{RF} - \omega_r$$

$$\text{or } \dot{\phi} = \omega_{RF} - ec^2 \frac{B}{E}$$

It is seen that with  $\omega_{RF}$  and  $B$  constant, it is not possible to have a synchronous particle with  $\phi = \text{cte}$  ( $\phi = 0$ ). The formula shows that  $\phi$  starts decreasing, reaches a minimum value ( $\phi = 0$ ) and then increases. Notice that the maximum tolerable  $\phi$  range of variation during acceleration is  $\pi$ .

Since phase and energy are related let us now consider their evolution during the acceleration cycle. A smooth approximation consists of assuming that the acceleration is made continuously, hence:

$$\dot{\phi} = \frac{\Delta\phi}{T_r/2} = \frac{\omega_r}{\pi} \Delta\phi$$

where  $\Delta\phi$  is the relative phase change at each half revolution,  $T_r$  being the revolution period. Hence one obtains:

$$\begin{aligned} \Delta\phi &= \frac{\pi}{\omega_r} \dot{\phi} = \pi \left( \frac{\omega_{RF}}{\omega_r} - 1 \right) \\ \Delta\phi &= \pi \left( \frac{\omega_{RF}E}{ec^2 B} - 1 \right) \end{aligned}$$

Differentiating the phase with respect to the energy, within the smooth approximation, leads to:

$$\frac{d\phi}{dE} \approx \frac{\Delta\phi}{\Delta E} = \frac{\pi}{e\hat{V} \sin \phi} \left( \frac{\omega_{RF} E}{ec^2 B} - 1 \right)$$

or

$$d(\cos \phi) = -\frac{\pi}{e\hat{V}} \left( \frac{\omega_{RF} E}{ec^2 B} - 1 \right) dE$$

Integrating both sides gives the energy-phase equation of the cyclotron:

$$\cos \phi = \cos \phi_0 - \frac{\pi}{e\hat{V}} \left[ \frac{\omega_{RF}}{2ec^2 B} (E^2 - E_0^2) - (E - E_0) \right]$$

where index 0 refers to the initial conditions:

$E_0$  = rest energy

$\phi_0$  = injection phase

It is interesting to represent on a graph the phase-energy evolution. Using the kinetic energy  $W = E - E_0$  as the new variable one gets:

$$\cos \phi = \cos \phi_0 + \frac{\pi}{e\hat{V}} \left( 1 - \frac{\omega_{RF}}{\omega_{r0}} \right) W - \frac{\pi}{2e\hat{V}E_0} \frac{\omega_{RF}}{\omega_{r0}} W^2$$

where  $\omega_{r0} = ec^2 B/E_0$  is the starting revolution frequency.

Figure 4 shows the variation of  $\cos \phi$  as a function of  $W$  for different initial conditions  $\phi_0$ . Plots which, during acceleration remain within the range  $-1 < \cos \phi < 1$ , only are meaningful from the stability point of view.

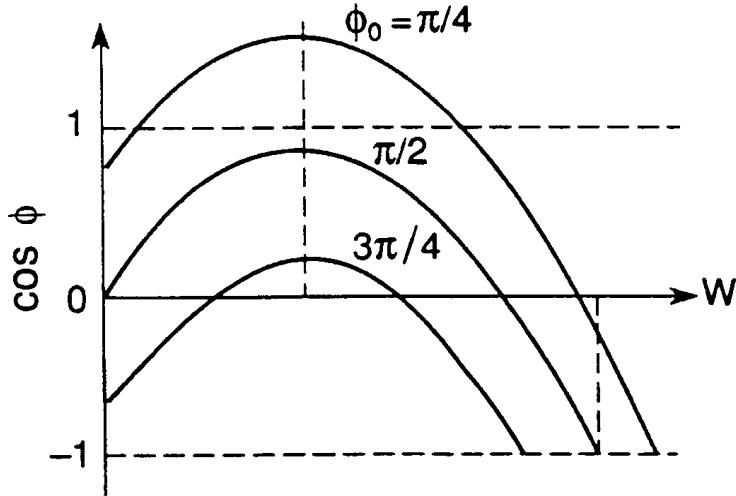


Fig. 4 Energy-phase diagram

Selected particles, according to their initial phase  $\phi_0$  can reach high energies, ;but the cyclotron is nevertheless limited in energy as seen in the figure. The machine parameters must be chosen to obtain the largest useful  $\phi_0$  range. This implies for instance  $\omega_{RF} < \omega_{r0}$  but it also depends very much on the particle species.

In the attempt to reach higher energies the concept of the synchrocyclotron has been quite important. Here the RF frequency is varied to keep the synchronous condition during acceleration. A new limitation however occurs due to the size of the magnet. A more efficient synchronous machine is the so-called synchrotron, which will be treated in more detail in the next sections.

### 1.3 Time varying electric field: the synchrotron case

A synchrotron is a circular accelerator where the nominal particle trajectory is kept at a constant physical radius by means of a magnetic field variation, as well as an RF frequency variation, to follow the energy variation.

Let's assume now that an RF cavity with an electric field component parallel to the particle trajectory is located at some azimuthal position of such a circular accelerator. In the cavity gap the electric field is supposed to have a sinusoidal time variation with angular frequency  $\omega_{RF}$ . Then one can write:

$$E(z, t) = E_1(z)E_2(t)$$

where  $E_1(z)$  is shown on Fig. 5 as a periodic function of period  $L = 2\pi R$ , while  $E_2(t)$  is of the form:

$$E_2(t) = E_0 \sin \left( \int_{t_0}^t \omega_{RF} dt + \phi_0 \right)$$

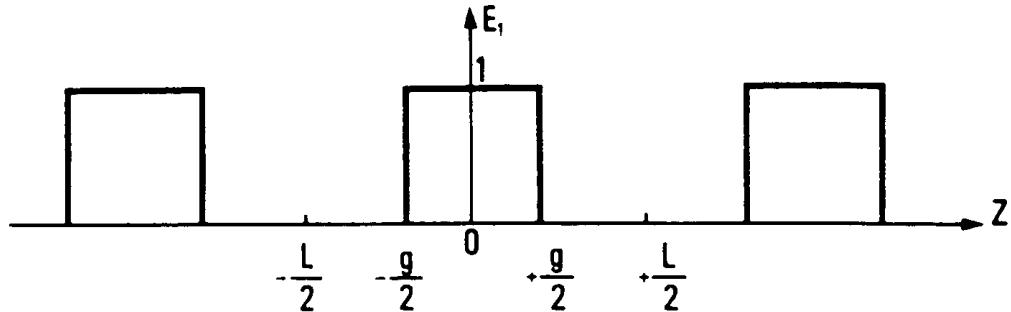


Fig. 5 RF field envelope along the circumference

The particle position is given by:

$$z = z_0 + \int_{t_0}^t v dt$$

and the RF frequency is chosen to be an integer multiple of the revolution frequency:

$$\omega_{RF} = h\omega_r = h \frac{2\pi\nu}{L}$$

where  $v$  is considered here as the average particle velocity over one turn.

The periodic function  $E_1(z)$  can be expanded in a Fourier series:

$$E_1(z) = A_0 + \sum_{n=1}^{\infty} A_n \cos\left(n \frac{2\pi z}{L}\right)$$

with

$$A_0 = \frac{g}{L}$$

$$A_n = \frac{2}{\pi n} \sin \frac{\pi n g}{L}$$

The integral of the force over many revolutions shows that all the trigonometric terms in the expansion will give no effective contribution to the energy gain, apart from a single term which corresponds to  $n = h$ . Then the average energy gain per turn can be written:

$$W = e \int_{-g/2}^{+g/2} E_1(z) E_2(t) dz = e E_0 g \sin \left( \int_{t_0}^t \omega_{RF} dt + \phi_0 \right)$$

with

$$\psi_0 = \phi_0 - \frac{2\pi h z_0}{L}$$

$$W = e E_0 \int_{-L/2}^{+L/2} A_h \cos^2 \frac{2\pi h z}{L} \sin \psi_0 \cdot dz$$

$$W = \frac{e E_0 A_h L}{2} \sin \psi_0 = \frac{e E_0 L}{\pi h} \sin \frac{\pi h g}{L} \sin \psi_0$$

$$W = e E_0 g \frac{\sin \frac{\pi h g}{L}}{\frac{\pi h g}{L}} \sin \psi_0 \cong e E_0 g \sin \psi_0 \quad \text{for } g \text{ small enough}$$

The energy gain per turn can also be expressed as:

$$W = e \int_{-g/2}^{+g/2} E_1(z) E_2(t) dz = e E_0 g \sin \left( \int_{t_0}^t \omega_{RF} dt + \phi_0 \right)$$

and since the gap center is at  $z = 0$  one has:

$$W = e E_0 g \sin \psi_0 = \hat{V} \sin \psi_0$$

showing that  $\psi_0$  represents the RF phase seen by the particle when crossing the gap.

From above it is seen that the force acting on the particle can be considered as an average force, continuously acting on the particle all over the circumference, provided the initial phase of the particle entering the gap is maintained constant. In other words the effect is equivalent to a force given by a travelling wave propagating at the same velocity as the particle velocity.

Since only one harmonic of the Fourier expansion is acting effectively, one can write the equivalent field as:

$$E = E_0 A_h \cos \frac{2\pi h z}{L} \sin \left( \int \omega_{RF} dt + \phi_0 \right)$$

$$E = \frac{E_o}{2} A_h \left\{ \sin \left( \int \omega_{RF} dt + \phi_0 - \frac{2\pi h z}{L} \right) + \sin \left( \int \omega_{RF} dt + \phi_0 + \frac{2\pi h z}{L} \right) \right\}$$

where the first term in the bracket represents a travelling wave with wavelength  $\lambda = L/h$  and phase velocity equal to the particle velocity, while the second term in the bracket gives no effect on a time average.

## 2. DISPERSION EFFECTS DUE TO THE GUIDE FIELD IN A SYNCHROTRON

### 2.1 Momentum compaction

By definition the momentum compaction  $\alpha$  is the constant parameter which relates the variation of the orbit length of a particle, in a circular accelerator, to the momentum deviation (note that the nominal closed orbit has been defined for the nominal energy  $E$ ).

$$\alpha = \frac{p}{L} \frac{dL}{dp} = \frac{p}{R} \frac{dR}{dp}$$

where  $p$ ,  $L$  and  $R$  are respectively the particle momentum, the nominal circumference and the physical radius ( $2\pi R = L$ ).

One has:

$$E = \gamma m_0 c^2 = (1 - \beta^2)^{-1/2} m_0 c^2$$

$$p = mv = \gamma \beta \frac{m_0 c^2}{c} = \beta (1 - \beta^2)^{-1/2} \frac{m_0 c^2}{c}$$

$$E = \frac{pc}{\beta}$$

and by logarithmic differentiation one gets:

$$\frac{dE}{E} = \frac{dp}{p} - \frac{d\beta}{\beta}$$

$$\frac{dp}{p} = \frac{d\beta}{\beta} + \frac{1}{2} \frac{2\beta d\beta}{1 - \beta^2} = \frac{d\beta}{\beta} \left( 1 + \frac{\beta^2}{1 - \beta^2} \right) = \frac{d\beta}{\beta} (1 - \beta^2)^{-1}$$

$$\frac{dE}{E} = \frac{dp}{p} - (1 - \beta^2) \frac{dp}{p} = \beta^2 \frac{dp}{p}$$

which leads to:

$$\frac{E}{R} \frac{dR}{dE} = \frac{\alpha}{\beta^2}$$

The average magnetic field along the nominal closed orbit is given by:

$$\begin{aligned}\langle B_{c.o.} \rangle &= \frac{1}{2\pi R} \oint B_{c.o.} ds \\ &= \frac{p/e}{2\pi R} \oint \frac{ds}{\rho} = \frac{p/e}{R} \quad (\rho = \text{bending radius}) \\ \langle B_{c.o.} \rangle R &= p/e\end{aligned}$$

and by differentiating

$$\begin{aligned}\frac{d\langle B_{c.o.} \rangle}{\langle B_{c.o.} \rangle} + \frac{dR}{R} &= \frac{dp}{p} \\ \frac{p}{\langle B_{c.o.} \rangle} \frac{d\langle B_{c.o.} \rangle}{dp} &= 1 - \alpha\end{aligned}$$

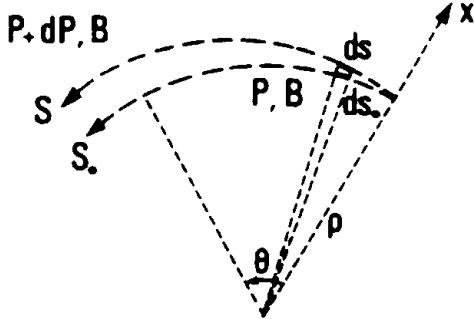
or

$$\frac{R}{\langle B_{c.o.} \rangle} \frac{d\langle B_{c.o.} \rangle}{dR} = \frac{1 - \alpha}{\alpha}$$

*Exercise: calculation of  $\alpha$*

Consider an isomagnetic guide field where all the bending magnets have the same curvature  $1/\rho$  and are separated by straight sections ( $1/\rho = 0$ ) which can include quadrupole magnets. To first-order approximation only the curved part of the orbit in the bendings contributes to a change of the length of the trajectory with momentum.

In a bending magnet one has:



$$\begin{aligned}ds_0 &= \rho d\theta \\ ds &= (\rho + x) d\theta \\ \frac{ds - ds_0}{ds_0} &= \frac{d\ell}{ds_0} = \frac{x}{\rho} = \frac{dx_{c.o.}}{\rho}\end{aligned}$$

The radial change in closed orbit with momentum is given by:

$$B\rho = P/e$$

$$D_x = \frac{dx_{c.o.}}{dp/p}$$

where  $D_x$  is the dispersion function (or local momentum compaction factor). A summation of all these small changes in the orbit length will give the change in the circumference.

$$\begin{aligned}\int d\ell &= dL = 2\pi dR \\ dR &= \frac{1}{2\pi} \int d\ell = \frac{1}{2\pi\rho_m} \int x ds_0 = \langle x \rangle_m\end{aligned}$$

where the subscript  $m$  means that the integral has to be calculated in the magnets only where  $1/\rho \neq 0$ . Finally one gets:

$$\alpha = \frac{p}{R} \frac{dR}{dp} = \frac{p}{R} \frac{\langle x \rangle_m}{dp} = \frac{\langle D_x \rangle_m}{R}$$

It can be shown that in smooth machines

$$\alpha \cong 1/Q_x^2$$

where  $Q_x$  is the radial wavenumber. In most cases however  $\langle D_x \rangle_m$  has to be numerically computed from lattice programmes.

## 2.2 Revolution frequency versus momentum

If  $f_r$  is the nominal revolution frequency corresponding to the nominal energy  $E$  (or momentum  $p$ ), the parameter  $\eta$  will be defined as follows:

$$\eta = \frac{p}{f_r} \frac{df_r}{dp}$$

Since

$$f_r = \frac{\beta c}{2\pi R}$$

one gets

$$\begin{aligned} \frac{df_r}{f_r} &= \frac{d\beta}{\beta} - \frac{dR}{R} = (1 - \beta^2) \frac{dp}{p} - \alpha \frac{dp}{p} \\ \frac{df_r}{f_r} &= \left( \frac{1}{\gamma^2} - \alpha \right) \frac{dp}{p} \end{aligned}$$

and hence

$$\eta = \frac{1}{\gamma^2} - \alpha$$

For an electron machine  $\eta \cong -\alpha$

## 2.3 Transition energy

The transition energy  $\gamma_{tr}$  is the energy which corresponds to  $\eta = 0$

$$\begin{aligned} \frac{1}{\gamma_{tr}^2} &= \alpha \\ \gamma_{tr} &= \sqrt{1/\alpha} \cong Q_x \end{aligned}$$

For small machines  $Q_x$  is of the order of a few units, while for very large machines it can approach 100. Hence,  $\gamma_{tr}$  will be in the range of 1 to 100 which is of interest only for proton machines because for electron machines  $\gamma \gg \gamma_{tr}$ .

Indeed it is possible in electron storage rings to make  $\alpha$  very small by using special focussing to make the transverse emittance very small as required for instance by synchrotron radiation users. In that case it is necessary to look to higher order in  $dp/p$  to get correct

dispersion relations. There are specific problems in proton synchrotrons related to the crossing of the transition energy which can be found in another lecture.

### 3. SYNCHROTRON OSCILLATION IN ADIABATIC LIMIT

One will consider the acceleration of particles with a radio frequency (RF) electric field which has a resonant condition with the nominal revolution frequency, or at least approximately.

#### 3.1 Synchronous particle

The accelerating voltage across the gap of the RF cavity can be expressed as:

$$V = \hat{V} \sin \int_0^t \omega_{RF} dt' = \hat{V} \sin \phi(t)$$

where  $\hat{V}$  and  $\omega_{RF}$  are slowly varying functions of  $t$ , but can also be constant as for instance in storage rings.

The RF frequency  $f_{RF}$  is set up to be an integer multiple of the nominal revolution frequency  $f_r$

$$f_{RF} = hf_r$$

where  $h$  is called the harmonic number. In these conditions a particle which has the nominal energy and circulates on the nominal trajectory will always experience the same RF phase when passing the RF gap:

$$\phi(t) = \phi_s$$

It is called the synchronous phase and it is related to the synchronous particle.

During the acceleration in a synchrotron the energy of the synchronous particle varies and so does the revolution frequency (unless particles are ultra relativistic). Clearly, if one wants to keep the accelerated particle on the same trajectory ( $R = cte$ ) the magnetic field must vary with time:

$$\begin{aligned} e\langle B \rangle R &= p = m_0 c \beta \gamma = m_0 c \beta (1 - \beta^2)^{-1/2} \\ \frac{dp}{dt} &= eR \frac{d\langle B \rangle}{dt} = eR \langle \dot{B} \rangle \end{aligned}$$

The energy gain per turn for the synchronous particle is:

$$(\Delta p)_{turn} = eR \langle \dot{B} \rangle T_r$$

where  $T_r$  is the revolution period:

$$T_r = \frac{1}{f_r} = \frac{2\pi R}{\beta c}$$

Hence:

$$(\Delta p)_{turn} = \frac{2\pi e R^2}{\beta c} \langle \dot{B} \rangle = \frac{2\pi e \rho R}{\beta c} \dot{B}$$

Since:

$$\Delta(E^2) = \Delta(p^2 c^2)$$

one gets:

$$\Delta E = \beta c \Delta p$$

and

$$(\Delta E)_{turn} = 2\pi e \rho R \dot{B}$$

where the energy gain per turn is obtained from the RF cavity:

$$(\Delta E)_{turn} = e \hat{V} \sin \phi_s$$

*Exercise*

Assume a 10 GeV proton synchrotron where the magnetic field reaches 1.5 Tesla in one second, following a linear variation. For a 10 GeV proton kinetic energy one has:

$$pc = (E^2 - m_0 c^2)^{1/2} = 10.9 \text{ GeV}$$

$$\beta = 0.996 ; \gamma = 11.7$$

hence

$$B\rho = P/e = 36.4 \text{ T.m}$$

and

$$\rho = 24 \text{ m } (R \approx 1.5\rho = 36 \text{ m})$$

As a result:

$$\hat{V} \geq 8.14 \text{ kVolts } (e \hat{V} \sin \phi_s \ll m_0 c^2)$$

Notice that in an electron synchrotron the particle radiates some energy per turn, and the amount of energy gain per turn must be greater than this loss in order to get an acceleration process.

### 3.2 Non-synchronous particle

In the following the parameters of the synchronous particle will be defined by subscript s. Any other particle will then be defined by its deviation from the synchronous one:

revolution frequency:  $f_r = f_{rs} + \Delta f_r$  (or  $\omega = \omega_s + \Delta\omega$ )

RF phase:  $\phi = \phi_s + \Delta\phi$

momentum:  $p = p_s + \Delta p$

energy:  $E = E_s + \Delta E$

$$\text{azimuthal angle: } \theta = \theta_s + \Delta\theta$$

The azimuthal angle is related to the azimuthal position by  $ds = R d\theta$ . Over one revolution this angle varies by  $2\pi$  while the RF phase varies by the quantity  $2\pi h$ . Hence, one has:

$$\Delta\phi = -h\Delta\theta$$

The - sign comes from the fact that a particle behind the synchronous particle ( $\Delta\theta < 0$ ) arrives later in the gap ( $\Delta t > 0$  and  $\Delta\phi > 0$ ). Moreover, since  $\theta = \int \omega dt$ , one has:

$$\Delta\omega = \frac{d}{dt}(\Delta\theta) = -\frac{1}{h} \frac{d}{dt}(\Delta\phi) = -\frac{1}{h} \frac{d\phi}{dt}$$

and from the definition of  $\eta$

$$\Delta p = -\frac{p_s}{h\eta\omega_s} \dot{\phi}$$

This can also be expressed in terms of energy:

$$\begin{aligned} \Delta E &= \frac{dE}{dp} \Delta p = v \Delta p = \omega R \Delta p \\ \frac{\Delta E}{\omega} &= -\frac{pR}{h\eta\omega} \dot{\phi} \end{aligned}$$

On each revolution the particle gains the energy:

$$(\Delta E)_{turn} = e\hat{V} \sin \phi$$

which corresponds to the momentum increment:

$$(\Delta p)_{turn} = \frac{e\hat{V}}{\omega R} \sin \phi$$

Dividing by the revolution period one gets the rate per second:

$$\dot{p} = \omega \frac{(\Delta p)_{turn}}{2\pi} = \frac{e\hat{V}}{2\pi R} \sin \phi$$

or

$$2\pi R \dot{p} = e\hat{V} \sin \phi$$

and for the synchronous particle:

$$2\pi R_s \dot{p}_s = e\hat{V} \sin \phi_s$$

By subtracting the two last expressions one gets:

$$2\pi \Delta(R\dot{p}) = e\hat{V}(\sin \phi - \sin \phi_s)$$

Expanding the left hand side to first order in  $\Delta R$  and  $\Delta\dot{p}$  gives:

$$\begin{aligned}
\Delta(R\dot{p}) &= R\dot{p} - R_s\dot{p}_s = (R_s + \Delta R)(\dot{p}_s + \Delta\dot{p}) - R_s\dot{p}_s \\
&\equiv \dot{p}_s\Delta R + R_s\Delta\dot{p} \\
&\equiv \dot{p}_s\left(\frac{dR}{dp}\right)_s \Delta p + R_s\Delta\dot{p} \\
&\equiv \dot{R}_s\Delta p + R_s\Delta\dot{p} = \frac{d}{dt}(R_s\Delta p) = \frac{d}{dt}\left(\frac{\Delta E}{\omega_s}\right)
\end{aligned}$$

The motion of any arbitrary particle in terms of deviations from the synchronous particle is then expressed by the following set of first order differential equations:

$$\begin{aligned}
\frac{dW}{dt} &= e\hat{V}(\sin \phi - \sin \phi_s) \\
\frac{d\phi}{dt} &= -\frac{1}{2\pi} \frac{h\eta\omega_s}{p_s R_s} W
\end{aligned}$$

where the new variable  $W = 2\pi R_s \Delta p = 2\pi \Delta E/\omega_s$  has been introduced.

It is worthwhile mentioning that the two variables  $\phi$ ,  $W$  are canonical since the equations of motion can be derived from a Hamiltonian  $H(\phi, W, t)$ :

$$\begin{aligned}
\frac{d\phi}{dt} &= \frac{\partial H}{\partial W} \\
\frac{dW}{dt} &= -\frac{\partial H}{\partial \phi}
\end{aligned}$$

with:

$$H(\phi, W, t) = e\hat{V}[\cos\phi - \cos\phi_s + (\phi - \phi_s)\sin\phi_s] - \frac{1}{4\pi} \frac{h\eta\omega_s}{R_s p_s} W^2$$

From the set of first-order equations of motion one can also derive a second-order differential equation for each variable. For instance the phase motion is given by:

$$\frac{d}{dt}\left[\frac{R_s p_s}{h\eta\omega_s} \frac{d\phi}{dt}\right] + \frac{e\hat{V}}{2\pi}(\sin\phi - \sin\phi_s) = 0$$

### 3.3 Small amplitude oscillations – phase stability

Let's consider the case where the parameters  $R_s$ ,  $p_s$ ,  $\eta$ ,  $\omega_s$  and  $\hat{V}$  are constant or at least change very slowly with time as compared to the variable  $\Delta\phi = \phi - \phi_s$ . Hence one can write:

$$\ddot{\phi} + \frac{\Omega_s^2}{\cos\phi_s}(\sin\phi - \sin\phi_s) = 0$$

where:

$$\Omega_s^2 = \frac{e\hat{V}h\eta\omega_s \cos\phi_s}{2\pi R_s p_s}$$

Within the approximation  $\Delta\phi \ll 1$  the equation of motion for small amplitudes becomes:

$$\ddot{\phi} + \Omega_s^2 \Delta\phi = 0$$

where  $\Omega_s$  now represents the synchrotron angular frequency. This quantity must be real in order to get a stable motion which means that  $\eta \cos \phi_s$  has to be positive. Stable synchrotron phase motion needs the following conditions to be satisfied:

$$\begin{aligned}\gamma < \gamma_{tr} \quad \eta > 0 \quad 0 < \phi_s < \frac{\pi}{2} \quad \sin \phi_s > 0 \\ \gamma > \gamma_{tr} \quad \eta < 0 \quad \frac{\pi}{2} < \phi_s < \pi \quad \sin \phi_s > 0\end{aligned}$$

having eliminated the cases where  $\sin \phi_s < 0$  which correspond to a deceleration.

At transition energy  $\eta$  vanishes,  $\Omega_s$  goes to zero and there is no more phase stability, at least within the first order approximation. During acceleration through transition energy, in a proton synchrotron, the RF phase must be switched rapidly from  $\phi_s$  to  $\pi - \phi_s$  in order to maintain stability above transition.

In the case of electron machines, either synchrotrons or storage rings, where the particle velocity is practically constant and equal to  $c$ , one has the following simplifications:

$$\begin{aligned}\omega_s = c/R_s \quad p_s = E_s/c \quad \eta \equiv -\alpha \\ \Omega_s = \frac{c}{R_s} \left\{ -\frac{h\alpha}{2\pi} \frac{\cos \phi_s}{E_s} \frac{eV}{E_s} \right\}^{1/2}\end{aligned}$$

The synchrotron tune which is the number of synchrotron oscillations per turn is represented by the bracket:

$$Q_s = \frac{\Omega_s}{c/R_s}$$

Note In an electron machine the RF frequency does not change and this is also true for  $R_s$  and  $\omega_s$ .

### 3.4 Large amplitude oscillations – RF acceptance

Considering again the equation of motion:

$$\ddot{\phi} + \frac{\Omega_s^2}{\cos \phi_s} (\sin \phi - \sin \phi_s) = 0$$

multiplying by  $\dot{\phi}$  and integrating lead to the invariant of the motion:

$$\frac{\dot{\phi}^2}{2} - \frac{\Omega_s^2}{\cos \phi_s} (\cos \phi + \phi \sin \phi_s) = cte$$

It is already known that around the stable synchronous phase  $\phi_s$  the small amplitude motions are pure harmonic oscillations which correspond to circles in the frame  $(\phi/\Omega_s, \phi)$ . For larger

amplitudes the circles will be distorted by the non-linearity of the motion but the curves will still close on themselves (Fig. 6). The extreme elongations of the oscillation correspond to  $\phi = 0$  and the constant of the motion can be expressed in terms of these values.

When  $\phi$  reaches the value  $\pi - \phi_s$  the factor  $(\sin \phi - \sin \phi_s)$  in the equation of motion becomes zero and for higher values of  $\phi$  the force is no more attractive so that the motion becomes unstable. Hence  $\pi - \phi_s$  is an extreme elongation corresponding to a stable motion. The corresponding curve in the  $(\phi/\Omega_s, \dot{\phi})$  space or in the  $(W, \phi)$  space is called the separatrix and the area delimited by this curve is called the RF bucket. The equation of the separatrix is:

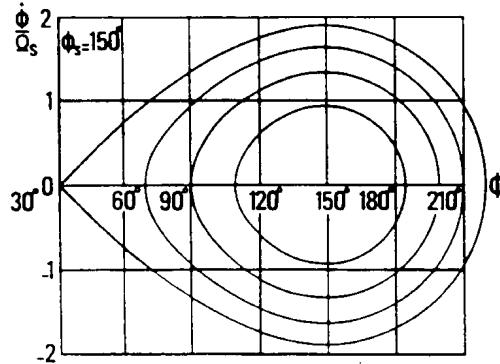


Fig. 6 Stable phase space trajectories

$$\frac{\dot{\phi}^2}{2} - \frac{\Omega_s^2}{\cos \phi_s} (\cos \phi + \phi \sin \phi_s) = -\frac{\Omega_s^2}{\cos \phi_s} [\cos(\pi - \phi_s) + (\pi - \phi_s) \sin \phi_s]$$

The second value  $\phi_m$ , for which  $\dot{\phi} = 0$ , is such that:

$$\cos \phi_m + \phi_m \sin \phi_s = \cos(\pi - \phi_s) + (\pi - \phi_s) \sin \phi_s$$

From the equation of motion it is also seen that  $\ddot{\phi}$  reaches an extremum when  $\ddot{\phi} = 0$  corresponding to  $\phi = \phi_s$ . Introducing this value in the equation of the separatrix gives the maximum stable values of  $\dot{\phi}$  and  $W$ :

$$\begin{aligned} \dot{\phi}_{max}^2 &= 2\Omega_s^2 [2 - (\pi - 2\phi_s) \operatorname{tg} \phi_s] \\ W_{max}^2 &= 2e\hat{V} [2 \cos \phi_s - (\pi - 2\phi_s) \sin \phi_s] \frac{2\pi p_s R_s}{h\eta\omega_s} \end{aligned}$$

or

$$\left( \frac{\Delta E}{E_s} \right)_{max} = \pm \beta \left\{ \frac{e\hat{V}}{\pi h\eta E_s} G(\phi_s) \right\}^{1/2}$$

This last expression is called the RF acceptance. The function  $G(\phi_s)$  is given by:

$$G(\phi_s) = [2 \cos \phi_s - (\pi - 2\phi_s) \sin \phi_s]$$

and varies from  $\pm 2$  to 0 when  $\sin \phi_s$  varies from 0 to 1.

The RF acceptance plays an important role when designing a machine, since it determines the capture efficiency at injection and the lifetime of stored beams. Outside the stable region plots of the trajectories (Fig. 7) show that particles get out of synchronism, their phase sliding along. Moreover the energy is continuously changing which means that the particles may get lost.

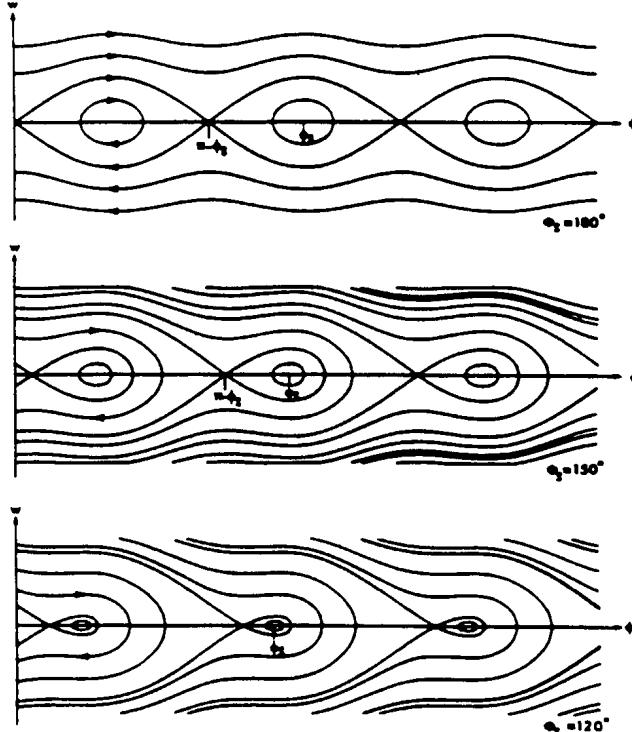


Fig. 7 Phase space trajectories for different  $\phi_s$

For any invariant of the motion there exists a relation between the maximum energy and the maximum phase deviations. However it is in general difficult to derive it analytically unless special assumptions are made. For instance in the case of small amplitude oscillations the invariant becomes simply:

$$\frac{\dot{\phi}^2}{2} + \Omega_s^2 \frac{\Delta\phi^2}{2} = cte$$

which leads to:

$$\Delta\phi_{max} = \frac{h\eta E_s}{p_s R_s \Omega_s} \left( \frac{\Delta E}{E_s} \right)_{max}$$

since  $\dot{\phi}_{max}$  is directly related to  $\Delta E_{max}$ .

In the case of ultra relativistic electrons this reduces to:

$$\Delta\phi_{max} = \frac{\alpha h}{Q_s} \left( \frac{\Delta E}{E_s} \right)_{max}$$

### 3.5 Potential energy function

The synchrotron motion is produced by a force field which can be derived from a scalar potential:

$$\frac{d^2\phi}{dt^2} = F(\phi)$$

$$F(\phi) = -\frac{\partial U}{\partial \phi}$$

$$U(\phi) = -\int F(\phi)d\phi = -\frac{\Omega_s^2}{\cos\phi_s}(\cos\phi + \phi \sin\phi_s)$$

The sum of the potential energy and the kinetic energy is a constant (the total energy):

$$\frac{\dot{\phi}^2}{2} + U(\phi) = U_0$$

The RF voltage as well as the corresponding potential energy function are shown on Fig. 8. The shape of the latter corresponds to the sum of a linear function and a sinusoidal one. An oscillation can only take place if the particle is trapped in the potential well which means that the total energy cannot exceed a certain value (dotted line) otherwise the particle will slide along the curve. Hence the maxima of the curve correspond to unstable equilibrium for the synchrotron motion.

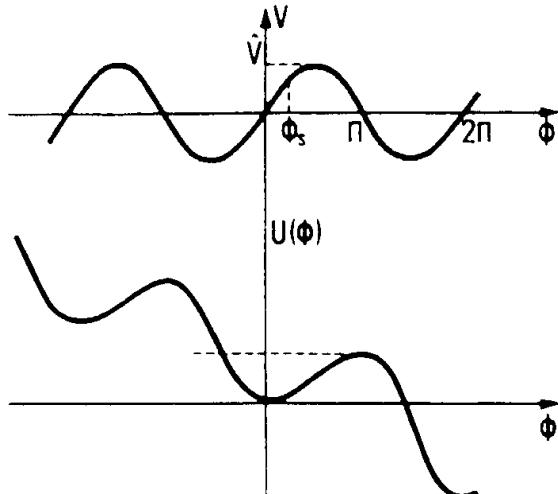


Fig. 8 Accelerating voltage and potential energy function

#### 4. ADIABATIC DAMPING OF SYNCHROTRON OSCILLATIONS

So far one has assumed that the parameters  $R_s$ ,  $p_s$ ,  $\omega_s$  and  $\hat{V}$  did not change appreciably at least over a time scale of one synchrotron period. However in a synchrotron these parameters will vary over a large range, even slowly, during an acceleration cycle. Then one needs to study the long term evolution of the motion under adiabatic changes of these parameters. This is possible with the help of the Boltzman-Ehrenfest adiabatic theorem which states that, if  $p$  and  $q$  are canonically conjugate variables of an oscillatory system with slowly changing parameters, then the action integral is constant:

$$I = \oint pdq = cte$$

the integral being taken over one period of oscillation. It has been already mentioned that the variables  $W$  and  $\phi$ , describing the synchrotron motion, were canonically conjugate. Hence applying the theorem leads to:

$$I = \oint W d\phi = cte$$

Consider the corresponding Hamiltonian of the motion and let's expand it to second order approximation to take care of small amplitude oscillations only

$$H(W, \phi, t) \approx -\frac{e\hat{V}}{2} \cos\phi_s \Delta\phi^2 - \frac{1}{4\pi} \frac{h\eta\omega_s}{R_s p_s} W^2$$

leading to harmonic solutions for the motions:

$$\begin{aligned} W &= \hat{W} \cos\Omega_s t \\ \Delta\phi &= \Delta\hat{\phi} \sin\Omega_s t \end{aligned}$$

Since:

$$\frac{d\phi}{dt} = \frac{\partial H}{\partial W} = -\frac{1}{2\pi} \frac{h\eta\omega_s}{R_s p_s} W$$

one gets the action integral:

$$\begin{aligned} I &= \oint W \frac{d\phi}{dt} dt \\ I &= -\frac{1}{2\pi} \frac{h\eta\omega_s}{R_s p_s} \oint W^2 dt \\ I &= -\frac{1}{2} \frac{h\eta\omega_s}{R_s p_s} \frac{\hat{W}^2}{\Omega_s} = cte \end{aligned}$$

where  $\hat{W}$  is the amplitude of the energy oscillation related to the amplitude of the phase displacement through:

$$\hat{W} = \frac{2\pi R_s p_s \Omega_s}{h\eta\omega_s} \Delta\hat{\phi}$$

So one gets:

$$\Delta\hat{\phi} \propto \left\{ \frac{\eta}{E_s R_s^2 \hat{V} \cos\phi_s} \right\}^{1/4}$$

Keeping all parameters constant except the energy which is ramping, the formula shows that the phase excursion  $\Delta\hat{\phi}$  is reduced as the one-fourth power of the energy.

It appears also that the product  $\hat{W} \cdot \Delta\hat{\phi}$  is invariant which means that the phase space area is invariant and Liouville's theorem still holds in adiabatic conditions. The phase space area is not damped, only the shape of the ellipse is modified.

From the previous treatment one also gets:

$$\Delta\hat{E} \propto \omega_s \left\{ \frac{E_s R_s^2 \hat{V} \cos\phi_s}{\eta} \right\}^{1/4}$$

$$\Delta\hat{R} \propto \frac{\alpha}{\omega_s} \left\{ \frac{\hat{V} \cos\phi_s}{E_s^3 R_s^2 \eta} \right\}^{1/4}$$

the last formula representing the orbit excursion due to the momentum deviation during the ramping.

The adiabatic damping can also be treated without the Hamiltonian formulation. Remembering the general equation for synchrotron oscillations limited to small phase deviations:

$$\frac{d}{dt} \left[ \frac{E_s}{h\alpha\omega_s^2} \frac{d\phi}{dt} \right] + \frac{e\hat{V} \cos\phi_s}{2\pi} \Delta\phi = 0 \quad (\text{for the case } \beta=1)$$

one can write it in the form:

$$\frac{E_s}{h\alpha\omega_s^2} \frac{d^2\phi}{dt^2} + \frac{\dot{E}_s}{h\alpha\omega_s^2} \frac{d\phi}{dt} + \frac{e\hat{V} \cos\phi_s}{2\pi} \Delta\phi = 0$$

or

$$\frac{d^2\phi}{dt^2} + \frac{\dot{E}_s}{E_s} \frac{d\phi}{dt} + \Omega_s^2 \Delta\phi = 0$$

where the second term represents a damping term. From the definition of  $\Omega_s$  one has:

$$\frac{\dot{E}_s}{E_s} = -2 \frac{\dot{\Omega}_s}{\Omega_s} .$$

To integrate this equation the procedure consists of choosing a solution similar to the one obtained without the additional damping term:

$$\Delta\phi = \Delta\hat{\phi} \sin \left( \int_0^t \Omega_s(\tau) d\tau + cte \right) = \Delta\hat{\phi}(t) \sin \psi(t)$$

and assuming  $\Delta\hat{\phi}$  and  $\dot{\Omega}_s$  are small first-order quantities (adiabatic limit). Putting this solution into the differential equation and neglecting all second-order terms gives:

$$\begin{aligned} & \left[ 2\Delta\hat{\phi}\dot{\Omega}_s - \Delta\hat{\phi}\dot{\Omega}_s \right] \cos\psi(t) = 0 \\ & 2\frac{\Delta\hat{\phi}}{\Delta\hat{\phi}} - \frac{\dot{\Omega}_s}{\Omega_s} = 0 \end{aligned}$$

and integrating leads to:

$$\Delta\hat{\phi} \propto \Omega_s^{1/2}$$

$$\Delta\hat{\phi} \propto E_s^{-1/4}$$

result which is similar to the one obtained previously.

## 5 . TRAPPING, MATCHING, ACCUMULATING AND ACCELERATING PROCESSES

Whether the circular accelerator is used as a synchrotron or a storage ring, the operation of the RF will be quite different.

### 5.1 Acceleration into a synchrotron

In that case, as mentioned before, the accelerating cycle is fast. Only a single injected pulse is accelerated. This injected pulse must be trapped in the RF buckets with a maximum efficiency which means that the RF acceptance hence the RF voltage has to be large enough, compatible with the energy spread of the transferred pulse. The RF frequency at injection must be such that it will fit with the bending field and the injection orbit. The synchronous phase is then set automatically at  $\phi_s = 0$  or  $\pi$  which means no average acceleration.

Matching means that the RF frequency and the RF voltage are adjusted such that the phase-space trajectories are homothetic to the contour of the injected bunch. If this was not the case the shape of the bunch would change during the synchrotron period and, for instance, the bunch length could become short giving rise to instabilities. Matching also requires careful adjustment of the injector to make it compatible with the possibilities of the synchrotron.

To start the acceleration it is necessary to move the synchronous phase so that the synchronous particle gets energy at each revolution from the RF cavities. This can be done by offsetting the magnetic field followed by a change of the RF frequency.

The synchronous particle is the one for which the revolution frequency satisfies

$$\omega_s = \frac{\omega_{RF}}{h}$$

and it follows a closed orbit for which the physical radius satisfies

$$\omega(B, R_s) = \omega_s$$

As mentioned previously the rate of change of momentum for the synchronous particle is:

$$\frac{dp_s}{dt} = eR_s \langle \dot{B} \rangle$$

Hence the RF frequency must follow:

$$\frac{f_{RF}(t)}{h} = \frac{1}{2\pi R_s} \beta c = \frac{1}{2\pi} \frac{e}{m(t)} \langle B(t) \rangle$$

$$\frac{f_{RF}(t)}{h} = \frac{1}{2\pi} \frac{ec^2}{E_s(t) R_s} \rho B(t)$$

Since

$$E_s^2 = (m_0 c^2)^2 + p_s^2 c^2$$

the RF frequency must follow the magnetic field variation with the following law:

$$\frac{f_{RF}(t)}{h} = \frac{c}{2\pi R_s} \left[ \frac{B(t)^2}{(m_0 c^2 / ecp)^2 + B(t)^2} \right]^{1/2}$$

This asymptotically tends towards  $f_r = c/2\pi R_s$  when  $B$  is getting large compared to  $m_0 c^2 / ecp$  corresponding to  $v \rightarrow c$  ( $pc \gg m_0 c^2$ ).

In practice the magnetic field can either follow an approximately linear law or a sinusoidal one

$$B(t) = \frac{\hat{B}}{2} (1 - \cos \omega t) = \hat{B} \sin^2 \frac{\omega}{2} t .$$

In the case of an electron synchrotron it is not necessary to vary the RF frequency because the particle velocity is very close to  $c$  and does not change with energy. However the electron loses energy in each revolution due to synchrotron radiation. Hence the synchronous particle is the one which arrives at the right phase  $\phi_s$  to compensate for both this energy lost and the field variation.

## 5.2 Accumulation into a storage ring

A storage ring is roughly a synchrotron operating at fixed energy. In some cases a very slow ramping can be done if the operating fixed energy differs from the injection one.

### 5.2.1 Electron storage rings

As for the electron synchrotron the energy lost has to be compensated. If the energy lost per turn is  $\delta E$ , then the synchronous phase is such that:

$$\delta E = e \hat{V} \sin \phi_s$$

which means that  $\phi_s$  will depend on the peak RF voltage  $\hat{V}$ . Moreover the energy lost per turn is a strong function of the operating energy:

$$\delta E_{[keV]} = 88.4 \frac{E_{[GeV]}^4}{\rho_{[m]}} .$$

If the storage ring operates between two energies, the maximum voltage will be determined by the upper energy and for the reason of beam lifetime due to particle diffusion through the separatrix,  $\sin \phi_s \leq 0.7$ . If one keeps at injection the same voltage as required at the highest energy then  $\sin \phi_s$  at injection can be very small leading to a large bucket acceptance.

Since the particle motion is damped around the synchronous particle the bunch length and energy spread become quite small leaving most of the bucket empty for a new injected pulse which will damp also and so on. This is the simple way of accumulating particles in an electron storage ring where Liouville's theorem does not hold any more due to non-conservative forces.

However accumulation of very high circulating currents in an electron storage ring may lead to typical instabilities related to coherent motion in the transverse phase space. Hence it is often desirable to keep  $Q$  constant when ramping the energy even slowly. Of course this will lower quite a lot the peak voltage required at injection and make the bucket smaller. Then the injector and the transport system to the storage ring have to be matched to the injection conditions.

### 5.2.2 Proton storage rings

Here the accumulation process is often called "stacking". It consists of trapping particles into buckets on a special orbit, called the injection orbit, close to the injection septum. Then the buckets are accelerated towards an inner orbit in the vacuum chamber. Such an acceleration is done with constant bending field, just by changing the RF frequency. Finally the RF voltage is switched off so that the particles will debunch. Hence, the RF is switched on again at the injection RF frequency to take care of new injected particles and the new buckets are accelerated to another stacking orbit close to the previous one. In doing so the previous injected particles will be slightly disturbed. The energy difference  $\Delta E_s$  between RF switch off of successive pulses is normally chosen to be approximately the bunch area divided by  $2\pi$  which corresponds to the energy width of an ideally debunched pulse. In this process the stacking efficiency suffers from the non-linearity of the motion in the neighbourhood of the separatrix.

## BIBLIOGRAPHY

M. Stanley Livingston, John P. Blewett. Particle Accelerators  
Mc Graw-Hill, Book Company, INC. N.Y 1962

E. Persico, E. Ferrari, S.E. Segre. Principles of Particle Accelerators, W.A. Benjamin, INC.  
N.Y 1968

H. Bruck, Accélérateurs circulaires de particules, Presses Universitaires de France, Paris 1966

B.W. Montague, RF acceleration, Proceedings of the first Course of the International School of  
Particle Accelerators, ERICE – CERN 77–13, 19 July 1977

M. Sands, The physics of electron storage rings, SLAC report 121 (1970). Also in  
Proceedings of the International School of Physics, B. Touschek, Editor (Enrico Fermi)  
Varenna 1969, Academic Press, N.Y. 1971

# GAMMA TRANSITION JUMP SCHEMES

T. Risselada

CERN, Geneva, Switzerland

## Abstract

Beam intensity loss and longitudinal emittance blow-up of high intensity proton beams at transition can be avoided if the speed of transition crossing is artificially increased by a so-called *gamma transition jump*, in which the value of  $\gamma_t$  is changed using fast pulsed quadrupoles. In order to improve the analysis of existing jump schemes and the preparation of new designs this lecture derives a simple expression for  $\Delta\gamma_t$  as a function only of the *local unperturbed* betatron and dispersion parameters at the jump quadrupoles and their integrated normalized gradients. The zero tune shift requirement has important consequences for the quadrupole layout and the number of families. The rules to be obeyed in the design of  $\gamma_t$  jump schemes are discussed, using the CERN PS and SPS machines as examples.

## 1 INTRODUCTION

In many medium and high energy (typically  $E > 5$  GeV) proton accelerators the beams have to cross *transition*, i.e. in the energy range of the machine there is a beam energy at which the derivative of the revolution frequency with respect to the momentum error crosses zero. The longitudinal beam behaviour in the neighbourhood of transition, and the radiofrequency manipulations required to accelerate the beams through transition, have been described in previous lectures [1]. The present lecture will deal with the problems caused to high intensity beams by longitudinal space charge forces, and more particularly with the most successful remedy, namely a fast  $\gamma_t$  jump. This solution has in the past twenty years allowed several machines to considerably raise the beam intensities that can be accelerated without intensity loss or emittance blow-up at transition.

The ratio between the beam energy at transition and the rest energy of the particles, called  $\gamma_t$  (gamma at transition), is independent of the particle mass. Its value depends only on the machine optics and geometry, and is close to the horizontal tune value in the case of a regular lattice. At constant magnetic field the velocity  $\beta c$ , the circumference  $C$ , and thus the revolution frequency  $f = \beta c/C$  of each particle depend on its relative momentum error  $\delta = \Delta p/p$ . The frequency spread in the beam is related [2] to its momentum spread by the parameter  $\eta$ :

$$\eta = \frac{1}{f_o} \frac{df}{d\delta} = \frac{1}{\gamma^2} - \frac{1}{C_o} \frac{dC}{d\delta} \quad (1)$$

where  $C_o$  and  $f_o$  are the circumference and the revolution frequency of a particle with nominal momentum on the reference orbit. At transition  $\eta = 0$  and the value of  $\gamma$  may be found from

$$\frac{1}{\gamma_t^2} = \frac{1}{C_o} \frac{dC}{d\delta} \quad (2)$$

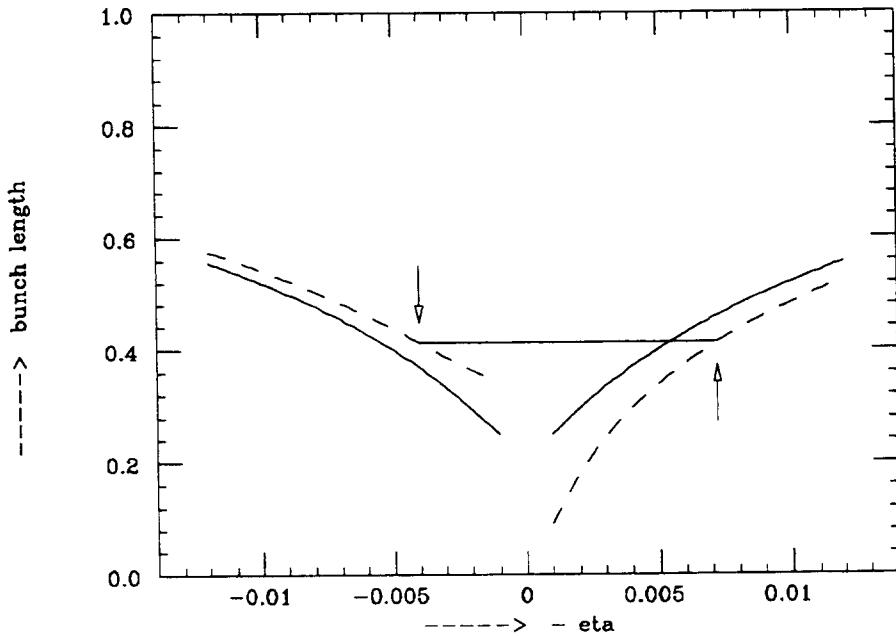


Figure 1: Schematic plot of equilibrium bunch length (in arbitrary units) before and after transition, without (solid) and with (dashed) space charge. The horizontal straight line between the arrows indicates the bunch length during a fast  $\gamma_t$  jump.

In the vicinity of transition  $\eta$  is approximately proportional to  $\gamma_t - \gamma$ :

$$\eta = \frac{1}{\gamma^2} - \frac{1}{\gamma_t^2} = \frac{\gamma_t^2 - \gamma^2}{\gamma^2 \gamma_t^2} \approx \left( \frac{2}{\gamma_t^3} \right) (\gamma_t - \gamma) \quad (3)$$

If the optics, and thus  $\gamma_t$ , is kept constant the speed at which the transition is crossed is determined only by the rate of acceleration  $d\gamma/dt$ . Beam intensity and emittances are in this case exposed to several intensity dependent space charge phenomena:

- The space charge effect on the bunch length changes sign at transition (figure 1), causing a sudden change of the equilibrium bunch length at a time when the motion in longitudinal phase space is slow [3, 4]. If the bunch shape oscillations resulting from this mismatch are not damped they will eventually result in filamentation and longitudinal emittance increase.
- Both the energy spread (Johnsen effect [5], intensity independent) and the space charge tune spread (Umstätter effect [4]) produce a spread in the  $\gamma_t$  value of the particles, which consequently do not cross transition simultaneously.
- Above transition the space charge forces make the bunches longitudinally unstable (negative mass instability [3, 4, 6, 7]). The growth rate of this instability increases rapidly with  $df/d\delta$  and thus with the time elapsed since transition. After some time it slows down and eventually the instability disappears when the Landau damping due to the increasing frequency spread in the bunches becomes large enough.

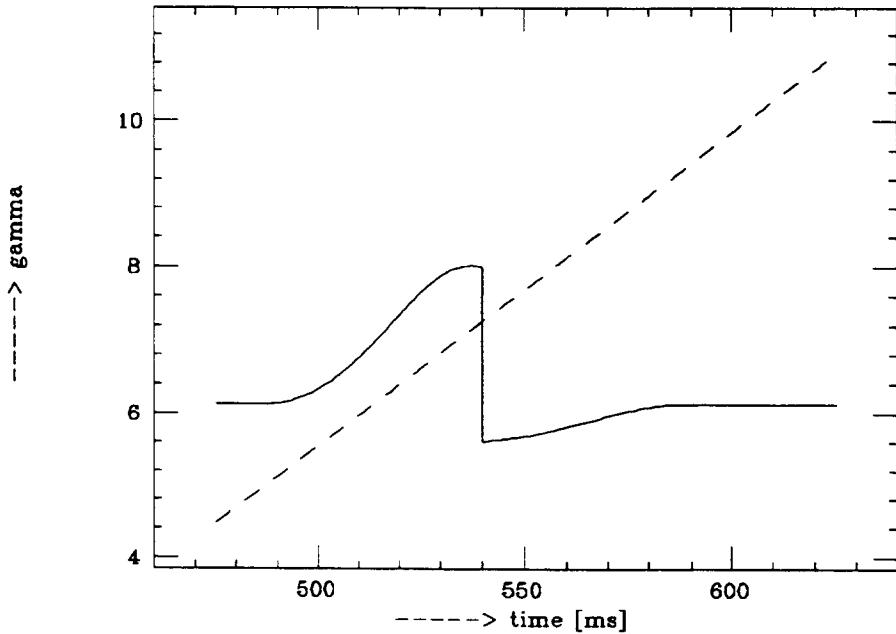


Figure 2:  $\gamma$  of beam (dashed) and  $\gamma_t$  (solid) at transition in the CERN PS. The jump is made in 0.4 ms.

In the design of accelerators it is attempted wherever possible to locate the transition energy outside the energy range of the machine, and some machines have even been built (for example the CERN LEAR) or proposed [8] with an *imaginary*  $\gamma_t$  value. On the other hand, the  $\gamma_t$  value of *existing* machines cannot be changed by a large amount without changing the ring geometry. If transition crossing cannot be avoided, the best solution to overcome the intensity limitations resulting from the above mentioned phenomena is an artificial increase of the transition crossing speed by means of fast pulsed quadrupoles. Such a  $\gamma_t$  jump scheme makes it possible to keep the beam at a safe distance from transition, except for the very short time during which the transition region is crossed (figure 2) at a speed increased by one or two orders of magnitude. The required jump amplitude and speed depend on the beam intensity [6, 9]:

- The distance between  $\gamma$  and  $\gamma_t$  just before and just after the jump should ideally be large enough for the space charge effect on the bunch length to be small compared to the unperturbed bunch length. With very high intensity beams this is not always possible, and it is therefore advantageous to use an asymmetric jump (figure 2), where the starting point of the jump is closer to transition than the end point. The equilibrium bunch length at the start and at the end (arrows in figure 1) can thus be made equal. The optimum asymmetry depends on the beam intensity.
- The negative mass unstable region just after transition should be crossed fast enough for the emittance blow-up to remain small. An asymmetric jump as mentioned above is better centered with respect to the unstable region than a symmetric jump.

For these reasons the optical design of a jump scheme, which is the subject of this lecture, should aim at a large  $\Delta\gamma_t$ , while keeping the maximum dispersion and  $\beta$  values below

reasonable values. The tune shift should be as small as possible. The use of a general purpose beam optics program does generally not allow an exhaustive study of a large number of possibilities (locations, polarities, families, etc.). Therefore we will first establish an analytical expression of the jump amplitude as a function of the quadrupole strengths, which will allow a preselection of a limited number of solutions. The performances (jump amplitude, maximum values of  $\beta$  and the dispersion, tune shift, etc.) of the selected candidate schemes may then be more precisely evaluated using a beam optics program.

## 2 EFFECT OF QUADRUPOLES ON $\gamma_t$

The modification of  $\gamma_t^{-2}$  due to the presence of non-nominal quadrupole gradients may be found, according to equation (2), by calculating the corresponding circumference increase of the closed orbit of a particle with momentum error  $\delta$ . To this end we consider a machine containing a set of  $N$  special quadrupoles which are not used in the nominal optics configuration, but may be powered to modify the horizontal dispersion function, for example at transition crossing. The nominal (unperturbed) machine has central orbit momentum  $p$ , magnetic rigidity of the beam  $B\rho = p/e$ , horizontal tune  $Q$ , and the following parameters valid at the special quadrupoles:

- $s_i$  the longitudinal position at quadrupole  $i \quad i = 1, N$
- $D_i$  the nominal horizontal dispersion
- $\beta_i$  the nominal horizontal beta
- $\mu_i$  the nominal horizontal betatron phase

These parameters are here assumed to be independent of  $\delta$ . The machine optics may be modified by powering the special quadrupoles (treated here as thin lenses), according to the parameters:

- $K_i$  the strength  $\int Gds/B\rho$  (integrated normalized gradient) of quadrupole  $i$
- $D_i^*$  the modified dispersion at quadrupole  $i$  in presence of strengths  $K_1, \dots, K_N$
- $\beta_i^*$  the modified horizontal beta

The nominal optics is characterized by the Courant - Snyder 'closed orbit distortion' function written here as

$$m(s_1, s) = \frac{\sqrt{\beta(s_1) \beta(s)}}{2 \sin(\pi Q)} \cos(\pi Q - |\mu(s_1) - \mu(s)|) \quad (4)$$

which is the effect of a unit kick at location  $s_1$  on the beam position at location  $s$  (figure 3). The horizontal closed orbit change due to a set of  $N$  non-nominal short horizontal orbit kicks  $\theta_i$  located at  $s_i$  may be written:

$$\Delta x(s) = \sum_{i=1}^N \theta_i m(s_i, s) \quad (5)$$

if  $\theta = \int Bds/(B\rho)$  is taken positive for a kick towards more positive  $x$  values. The same function  $m$  relates the nominal dispersion to the local bending radius  $\rho(s)$  in the bending magnets [10]:

$$D_i = \int_0^C \frac{m(s, s_i)}{\rho(s)} ds \quad (6)$$

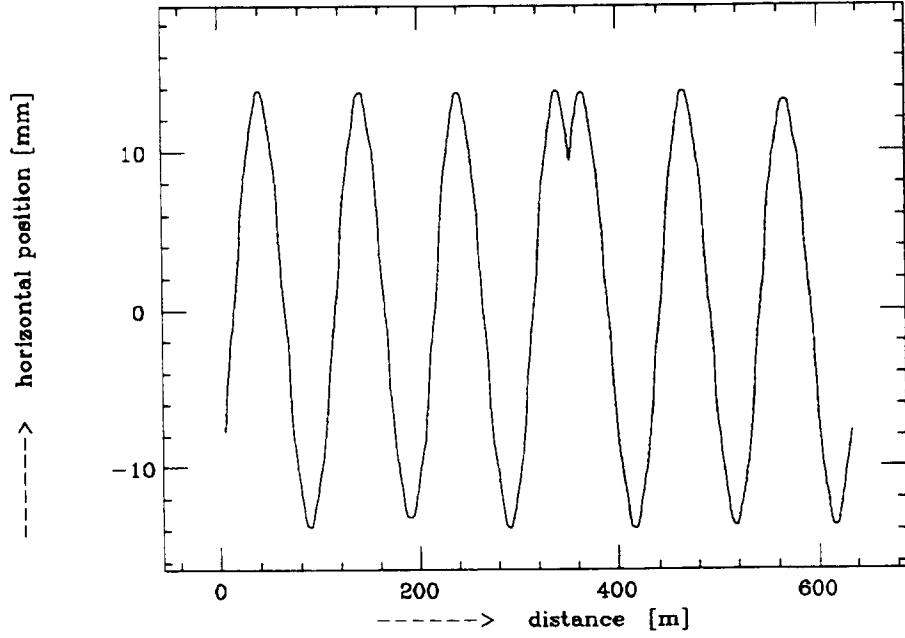


Figure 3: Closed orbit in the presence of a single kick (sampled at locations with equal  $\beta$ )

The parameter  $s$  denotes the projection of the longitudinal position onto the central orbit. To first order in  $x$  the ratio between a path length on an orbit  $x(s)$  and its projection is equal to  $1 + x(s)/\rho(s)$  in the bending magnets and equal to 1 elsewhere [10]. The circumference increase of an orbit in the presence of kicks  $\theta_i$ , calculated from the path length increase in the bending magnets, may thus be written

$$\Delta C = \int_0^C \frac{\Delta x(s)}{\rho(s)} ds \quad (7)$$

Using (5) and (6) this may be related to the dispersion:

$$\Delta C = \sum_{i=1}^N \theta_i \int_0^C \frac{m(s_i, s)}{\rho(s)} ds = \sum_{i=1}^N \theta_i D_i \quad (8)$$

Since we are only looking at the properties of one orbit at a time it is irrelevant whether a kick is caused by a *dipole* field or by a *quadrupole* field, provided both have the same field value at the radial position of the orbit. The special quadrupoles may thus be treated as kicks, and their effect on the orbit and the circumference may be calculated using the dispersion,  $\beta$  values and phases of the *nominal* optics. On the other hand, the value of the orbit kick resulting from the strength  $K_i$  of (thin) quadrupole number  $i$  depends on the orbit position in the quadrupole, which is equal to the *perturbed* local dispersion  $D_i^*$  multiplied by the momentum error  $\delta$ :

$$\theta_i = -x(s_i) K_i = -\delta D_i^* K_i \quad (9)$$

if  $K$  is taken positive for a horizontally focussing quadrupole. According to (8) the effect of a set of  $N$  quadrupole strengths  $K_i$  on the circumference is then

$$\Delta C = -\delta \sum_{i=1}^N K_i D_i^* D_i \quad (10)$$

The dispersion functions are here assumed to be independent of  $\delta$ . Differentiation with respect to  $\delta$  yields the  $\gamma_t^{-2}$  change

$$C_o \Delta(\gamma_t^{-2}) = \Delta \left( \frac{dC}{d\delta} \right) = - \sum_{i=1}^N K_i D_i^* D_i \quad (11)$$

which thus depends only on the *local* values of the perturbed and unperturbed dispersion functions at the  $N$  quadrupoles.

The nominal and perturbed dispersion functions are related by the closed orbit distortion function  $m$ . The closed orbit  $x(s_j)$  at quadrupole  $j$  of a particle with relative momentum error  $\delta$  in the presence of the  $N$  special quadrupole strengths  $K_i$  is:

$$x(s_j) = \delta D_j - \sum_{i=1}^N \delta K_i D_i^* m(s_i, s_j) \quad (12)$$

Differentiation with respect to  $\delta$  yields

$$\vec{D}^* = \vec{D} + M \cdot \vec{D}^* \quad (13)$$

where vector  $\vec{D}$  is defined as  $\vec{D} = (D_1, D_2, \dots, D_N)$  and matrix  $M$  as:

$$M_{ij} = -K_j \quad m_{ij} = -K_j \quad m(s_i, s_j) \quad i, j = 1, N \quad (14)$$

valid for the nominal unperturbed machine. The modified dispersion values may be found by

$$\vec{D}^* = (1 - M)^{-1} \cdot \vec{D} = (1 + M + M^2 + \dots) \cdot \vec{D} \quad (15)$$

### 3 ZERO TUNE SHIFT AND NON-ZERO $\Delta(\gamma_t)$

Equations (11) and (15) have a general validity and may be applied to any  $\gamma_t$  modification system, DC or pulsed. However, a pulsed jump scheme for transition crossing has to modify  $\gamma_t$  temporarily without creating additional instabilities or beam loss, and the optical design must therefore meet three requirements simultaneously:

- produce a large  $\gamma_t$  change
- leave the betatron tunes unchanged
- keep  $D_{max}$  and  $\beta_{max}$  below reasonable values

The first two requirements may a priori seem incompatible, as  $\gamma_t$  is usually close to the horizontal tune. However, they may be met by locating the quadrupoles in such a way that they act strongly on the dispersion and little on  $\beta$ , since  $\gamma_t$  is related to the dispersion and  $Q$  to the  $\beta$  function. This is possible by exploiting the difference in periodicity of the functions  $\Delta D$  and  $\Delta \beta$ , as will be shown below.

### 3.1 How to obtain $\Delta Q = 0$

To first order in  $K_i$  the tune shift is given by [10]

$$\Delta Q = \frac{1}{4\pi} \sum_{i=1}^N \beta_i K_i \quad (16)$$

Thus  $\sum \beta_i K_i = 0$  is a first requirement to keep the horizontal tune constant. An equivalent condition should be respected in the vertical plane. Unfortunately, if no special precautions are taken the  $\beta$  functions are modified by the jump quadrupoles, and the expression for the tune shift contains higher order terms in  $K_i$ . However, it is possible to keep the tunes constant during the entire transition process if the quadrupole locations are chosen with care. This is illustrated by a short review of early  $\gamma_t$  jump schemes:

- A non-zero  $\Delta Q$  was tolerated at the CERN PS between 1969 and 1973 with the so-called Q-jump scheme [11, 12], using a set of 6 quadrupoles more or less regularly spaced around the ring with *identical* strengths and polarities. A tune change of 0.25 was required to obtain a  $\Delta \gamma_t$  of 0.3.
- In 1970 a scheme was proposed for the FNAL Booster using 12 regularly spaced quadrupoles with equal strengths but *alternating* polarities [13], yielding  $\Delta \gamma_t = 1$  at the expense of only  $\Delta Q = 0.1$ .
- At the PS a large  $\gamma_t$  jump of 2.7 *without* tune shift is obtained [6] since 1973 by grouping 16 quadrupoles together in doublets  $(i, i+1)$  for  $i = 1, 3, \dots$  with  $K_{i+1} = -K_i$  and  $\mu_{i+1} - \mu_i = \pi$ .

The last example is the only one of the three cases where the  $\beta$  values are kept constant at the jump quadrupoles. This can be obtained in any machine using the simple and yet extremely flexible solution of grouping the quadrupoles together as doublets in such a way that for each *individual* doublet the betatron transfer matrix (from the entrance of the first lens to the exit of the second) does not depend on the excitation level. It can be shown that this is equivalent to the two conditions

$$\beta_1 K_1 + \beta_2 K_2 = 0 \quad \sin(\mu_2 - \mu_1) = 0 \quad (17)$$

The second condition implies  $\Delta\mu = n\pi$  with integer  $n$ . Such doublets leave the  $\beta$  values strictly unchanged outside the doublet interval, and thus at *all* quadrupole locations in the machine. The doublets must not overlap. Any scheme consisting of such doublets will then yield  $\Delta Q = 0$  even for large  $K$  values. Due to the difference in periodicity of the  $\Delta D$  and  $\Delta \beta$  functions the effect on  $\Delta(\gamma_t^{-2})$  is zero or small for *even*  $n$ , and large for *odd*  $n$  values. In the following we will therefore assume a  $\pi$  phase advance in all doublets.

### 3.2 Consequences of the use of $\pi$ -doublets

A  $\pi$  phase advance in a zero tune shift doublet eliminates all powers  $> 1$  of matrix  $M$  in equation (15). This may be shown as follows:

$$(M^2)_{ij} = \sum_k K_k m_{ik} K_j m_{kj} = K_j \sqrt{\beta_i \beta_j} \sum_k K_k \beta_k \cos_{i,k} \cos_{k,j} \quad (18)$$

where

$$\cos_{i,j} = \frac{1}{2 \sin(\pi Q)} \cos(\pi Q - |\mu(s_i) - \mu(s_j)|) \quad (19)$$

The contribution of doublet  $(k, k + 1)$  to each element  $(M^2)_{ij}$  is proportional to  $K_k \beta_k + K_{k+1} \beta_{k+1}$ , since  $\cos_{i,k} = -\cos_{i,k+1}$  and  $\cos_{k,j} = -\cos_{k+1,j}$ . The zero tune shift condition cancels this contribution and equation (15) has only 2 terms:

$$\vec{D}^* = (1 + M) \cdot \vec{D} \quad (20)$$

Equation (11) then reduces to:

$$C_o \Delta(\gamma_t^{-2}) = - \sum_{i,j=1}^N K_i (1 + M)_{ij} D_i D_j = - \sum_{i=1}^N K_i D_i^2 + \sum_{i,j=1}^N K_i K_j m_{ij} D_i D_j \quad (21)$$

valid for a  $\pi$ -doublet scheme with zero tune shift.

The effects of a  $\pi$ -doublet with  $\beta_1 K_1 + \beta_2 K_2 = 0$  may thus be summarized as follows:

- $Q$  is *unchanged*
- $\beta$  is left *unchanged* at all jump quadrupoles
- the dispersion change is *linear* in  $K$
- the jump amplitude has only *linear and quadratic* terms in  $K$

The maximum  $\beta$  value inside the doublet interval can easily be found in most cases. For example in a regular lattice where all horizontally focussing quadrupoles are located at  $\beta_i = \beta_{max}$  the maximum perturbed  $\beta^*$  inside doublet  $(i, i + 1)$  may be found [6] by

$$\frac{\beta_{max}^*}{\beta_{max}} = \frac{1 + \sin \psi}{1 - \sin \psi} \quad (22)$$

where  $\psi$  is defined by  $\tan(\psi) = |K_i| \beta_{max}/2$ . On the other hand, the dispersion function modification is generally non-local and the maximum dispersion value cannot be estimated a priori in a simple way.

### 3.3 Schemes with a constant normalized dispersion

In schemes where the horizontal normalized dispersion  $D/\sqrt{\beta}$  has the same value at all quadrupoles the  $\sum \beta_i K_i = 0$  requirement cancels the linear terms  $-\sum K_i D_i^2$  of equation (21), and  $\Delta(\gamma_t^{-2})$  will only have quadratic and higher order terms. If the scheme consists of  $\pi$ -doublets the higher order terms cancel and the  $\gamma_t$  change is given by the quadratic terms of (21):

$$\Delta(\gamma_t^{-2}) = \frac{1}{2C_o \sin(\pi Q)} \sum_{i,j=1}^N D_i^2 \beta_j K_i K_j \cos(\pi Q - |\mu_i - \mu_j|) \quad (23)$$

valid for a  $\pi$ -doublet scheme with zero tune shift and identical nominal normalized dispersion values  $D/\sqrt{\beta}$  at all  $N$  quadrupoles.

## 4 CELLS, SUPERPERIODS AND FAMILIES

As shown above, the expression for  $\Delta(\gamma_t^{-2})$  contains only first and second order terms in  $K$  in  $\pi$ -doublet schemes. In other schemes the higher order terms are usually small and  $\Delta(\gamma_t^{-2})$  is therefore in all cases dominated either by linear or by quadratic terms in  $K$ :

- In machines where the normalized dispersion function has large variations from one potential quadrupole location to another the quadrupoles of a doublet may be located at  $(D/\sqrt{\beta})_{max}$  and  $(D/\sqrt{\beta})_{min}$  respectively, in order to maximize the linear term  $-\sum K_i D_i^2$  of equation (21) which dominates  $\Delta(\gamma_t^{-2})$  in this case. In regular lattice structures these locations are separated by a  $\pi$  betatron phase advance.
- In machines with an *equal* normalized horizontal dispersion  $D/\sqrt{\beta}$  at all quadrupoles the value of  $\Delta(\gamma_t^{-2})$  depends quadratically on the quadrupole strengths  $K_i$  for zero tune shift schemes.

Most machines are made up of repetitive cells, each containing a small number of quadrupoles, for example (QF, QD). The vacuum chamber aperture is used in the most efficient way if the  $\beta$  and dispersion functions are matched to these cells, i.e. have the same periodicity. Free space availability and the preference for cell phase advances of  $45^\circ$ ,  $60^\circ$  or  $90^\circ$  require quadrupoles of a  $\pi$ -doublet to occupy *equivalent* locations in different cells, resulting in identical  $D$  and  $\beta$  values (example of the PS). On the other hand, the presence of straight sections with zero dispersion, as required in some machines, induces  $D/\sqrt{\beta}$  variations around the ring (example of the SPS).

### 4.1 Example of matched normalized dispersion: the CERN PS

At the CERN PS ( $Q = 6.25$ ,  $E_{min} = 2$  GeV,  $E_{max} = 26$  GeV,  $\gamma_t = 6.1$ ,  $d\gamma/dt = 40$  s $^{-1}$ )  $D$  and  $\beta$  have the same value at all potential jump quadrupole locations and consequently  $\Delta(\gamma_t^{-2})$  depends quadratically on the  $K$  values. This means that, with one power supply feeding all quadrupoles, it would only be possible to change  $\gamma_t$  in one direction, depending on the sign of  $\sin(\pi Q)$ . However in this case the maximum change of  $\gamma_t$ , and thus the maximum excursions of the dispersion and  $\beta$  functions with respect to their nominal values, would be much larger than with a *symmetric bipolar*  $\gamma_t$  jump providing the same jump amplitude. At the PS the quadrupole doublets are for this reason separated in two families, each connected to a separate power supply [6]. If family number 1 has strengths  $+K_1$  or  $-K_1$ , and family number 2 strengths  $+K_2$  or  $-K_2$ , equation (23) may be rewritten as a sum of three terms:

$$\Delta(\gamma_t^{-2}) = aK_1^2 + bK_1K_2 + cK_2^2 \quad (24)$$

The phases  $\mu_i$  of the quadrupole locations have to be chosen in such a way that coefficient  $b$  is large, and larger than  $a$  or  $c$ . A polarity change of either  $K_1$  or  $K_2$  then changes the sign of  $\Delta(\gamma_t^{-2})$ . At the PS both families are slowly pulsed with rise and fall times of about 20 ms. At the maximum of the current pulses the polarity of family 1 is reversed as fast as possible, which yields the  $\gamma_t$  curve shown in figure 2.

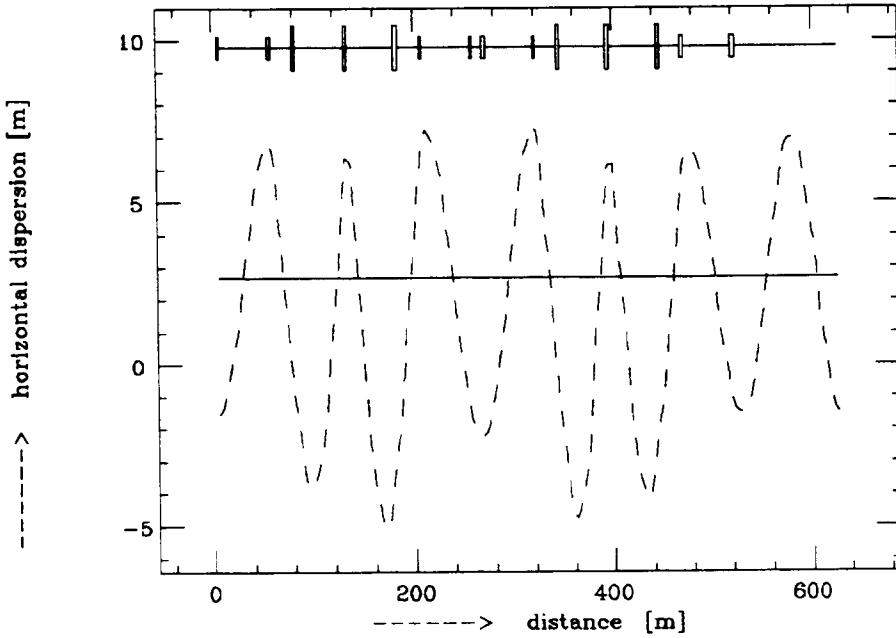


Figure 4: Unperturbed dispersion (solid) of the PS and dispersion just before the polarity change of the gamma jump (dashed), sampled at locations with equal  $\beta$ . Family 1 is shown as the smaller boxes, family 2 as the larger boxes.

The basic brick of the present PS scheme is

$$(+D_1) \quad \phi \quad (+D_2) \quad (25)$$

where  $+D_1$  denotes a doublet consisting of two quadrupoles at  $\pi$  phase advance, with strengths  $+K_1$  and  $-K_1$ . The  $45^\circ$  phase advance per cell in both planes makes it easy in this machine to locate the doublet quadrupoles at the required phase advance.  $D_1$  is separated from the second doublet  $D_2$  (strengths  $+K_2$  and  $-K_2$ ) by a space with  $\phi$  betatron phase advance, not containing jump quadrupoles. In order to keep  $D_{max}^*$  low it is advantageous to make the quadrupole layout symmetric. The simplest symmetric configuration is:

$$(+D_1) \quad \phi \quad (+D_2) \quad (-D_2) \quad \phi \quad (-D_1) \quad (26)$$

where  $-D$  has reversed polarities with respect to  $+D$ . The two central doublets together form a triplet with strengths  $+K_2$ ,  $-2K_2$ ,  $+K_2$ . This will have the same effect on  $\gamma_t^{-2}$  as a single doublet with strengths  $+2K_2$  and  $-2K_2$ , but the local maximum  $\beta$  value will be smaller, according to (22). Doublets or groups of doublets may be displaced by  $\pi$  phase advance if their polarities are reversed, without changing the  $\Delta(\gamma_t^{-2})$  value, since the  $\beta$  and dispersion values are identical at all considered quadrupole locations. The  $\gamma_t$  jump produced by the scheme depends thus on  $\phi$  modulo  $\pi$ .

For  $0 < \phi < \pi$  this configuration stretches over a phase advance between  $4\pi$  and  $6\pi$ , and with a tune of 6.25 there is space for installing two such configurations around the ring. A given jump amplitude may thus be obtained with smaller quadrupole strengths, resulting once more in smaller  $D_{max}^*$  and  $\beta_{max}^*$  values. The value of  $\phi$  and the phase

advance between the two configurations has to be chosen with care in order to optimize  $a$ ,  $b$  and  $c$ . With the PS cell phase advance of  $45^\circ$  only 16 independent cases have to be taken into consideration. We will here limit the discussion to two solutions which we call '*symmetric*' and '*antisymmetric*' respectively. At this stage the manipulations may more easily be understood by talking in terms of superperiods. With the above mentioned phase advances the machine may be divided either into two identical superperiods containing jump quadrupoles, or into two identical superperiods containing quadrupoles plus one transparent betatron wavelength without quadrupoles.

The symmetric scheme corresponds to two exactly identical superperiods, each having a tune of  $Q/2$ . The coefficients  $a$ ,  $b$  and  $c$  may be evaluated using equation (23) putting  $Q = 6.25/2$ . A large  $b$  value is obtained for  $\phi = \pi/2$ , but  $c$  is equally large (values in  $\text{m}^2$ ):

$$a = 0 \quad b = -7.2 \quad c = +8.7 \quad (27)$$

The jump is thus not bipolar, and the performance of this scheme is poor:

$$\Delta\gamma_t = 1.0 \quad D_{max}^* = 21.0 \text{ m} \quad (28)$$

However, with  $K_1 = -K_2$  the scheme provides a very low  $\gamma_t$  value of 4.3, which may be useful for other purposes [6].

The scheme may be made antisymmetric by either reversing the polarity of one of the two quadrupole configurations, or displacing one configuration over a  $\pi$  phase advance. The antisymmetric case thus corresponds to two identical superperiods with a tune of  $(Q - 1)/2$ , plus one 'transparent' betatron period without quadrupoles (figure 4) and yields with  $\phi = \pi/2$  (values in  $\text{m}^2$ ):

$$a = 0 \quad b = -7.2 \quad c = -1.5 \quad (29)$$

The last solution was adopted for the jump scheme of the PS as it provides a large bipolar  $\gamma_t$  jump. The non-zero  $c$  coefficient makes  $\Delta(\gamma_t^{-2})$  slightly asymmetric:  $\gamma_t$  increases first slowly from 6.1 to 8.2, jumps down to 5.6, and rises back slowly to 6.1 again (figure 2). The maximum dispersion and  $\beta$  values just before or just after the jump are 9 m and 54 m respectively, to be compared with the nominal values of 3 m and 22 m (figure 4). The corresponding quadrupole strengths are

$$K_1 = 0.023 \text{ m}^{-1} \quad K_2 = 0.050 \text{ m}^{-1} \quad (30)$$

The maximum current in the quadrupoles is 450 and 250 A respectively, and the polarity of  $K_1$  is reversed in 0.4 ms at transition, using special low inductance quadrupoles. This yields a transition crossing rate of  $-d\gamma_t/dt = 6000 \text{ s}^{-1}$ , which is 150 times larger than  $d\gamma/dt$ .

#### 4.2 Example of mismatched normalized dispersion: the CERN SPS

Not all machines have equal  $D/\sqrt{\beta}$  values at all quadrupoles. As a counter-example we will briefly discuss a possible  $\gamma_t$  jump scheme for the SPS ( $Q = 26.6$ ,  $E_{min} = 14 \text{ GeV}$ ,  $E_{max} = 450 \text{ GeV}$ ,  $\gamma_t = 23.2$ ,  $d\gamma/dt = 130 \text{ s}^{-1}$ ) where presently high intensity proton beams cross transition without a  $\gamma_t$  jump. This machine has a regular lattice

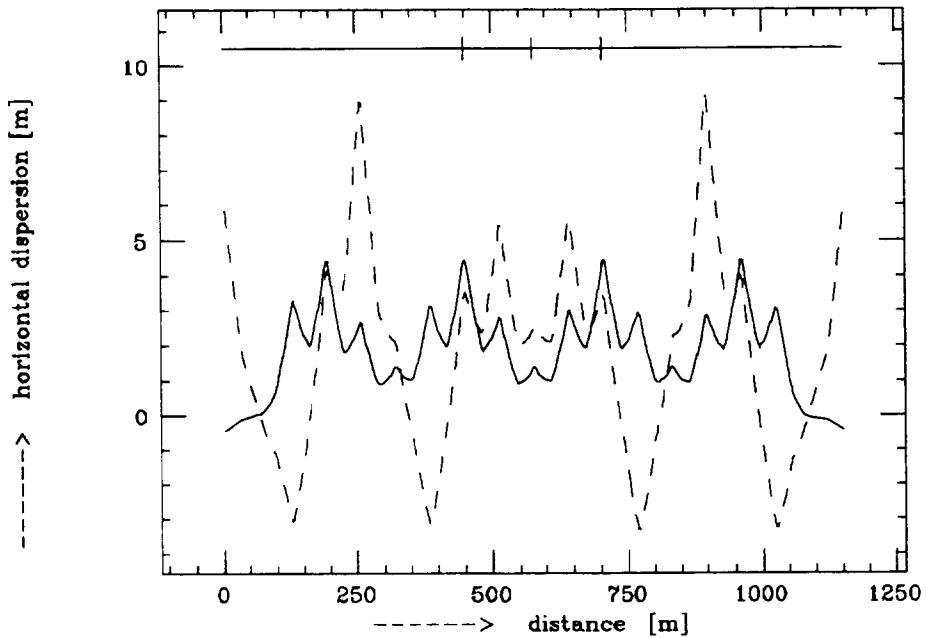


Figure 5: A possible  $\gamma_t$  jump scheme for the SPS: unperturbed (solid) and perturbed (dashed) dispersion in one superperiod (SPS/6), with proposed jump quadrupole locations.

structure in the arcs, but the matching of the dispersion function is destroyed by the presence of 6 long straight sections with zero dispersion: the  $\beta$  values are equal at all F quadrupoles, while the dispersion oscillates between 1 and 4.5 m (figure 5). Since the cell phase advance is nearly  $90^\circ$  it is easy to form doublets yielding  $\Delta Q \approx 0$ .

With such a dispersion modulation the expression (21) for  $\Delta(\gamma_t^{-2})$  may contain a large linear term  $C_o^{-1} \sum K_i D_i^2$  and thus one quadrupole family is sufficient to produce a bipolar  $\gamma$  jump. In the regular SPS arc lattice  $D_{max}$  and  $D_{min}$  are separated by a  $\pi$  betatron phase advance, and doublet quadrupoles located there produce a large  $\Delta\gamma_t$ . With one triplet (i.e. two contiguous doublets) per superperiod, centered in the arc,  $K$  values of only + and -  $0.010 \text{ m}^{-1}$  are required to obtain a  $\Delta\gamma_t$  of 3.6.  $D_{max}$  increases from 4.5 to 9 m, and  $\beta_{max}$  from 104 to 226 m.

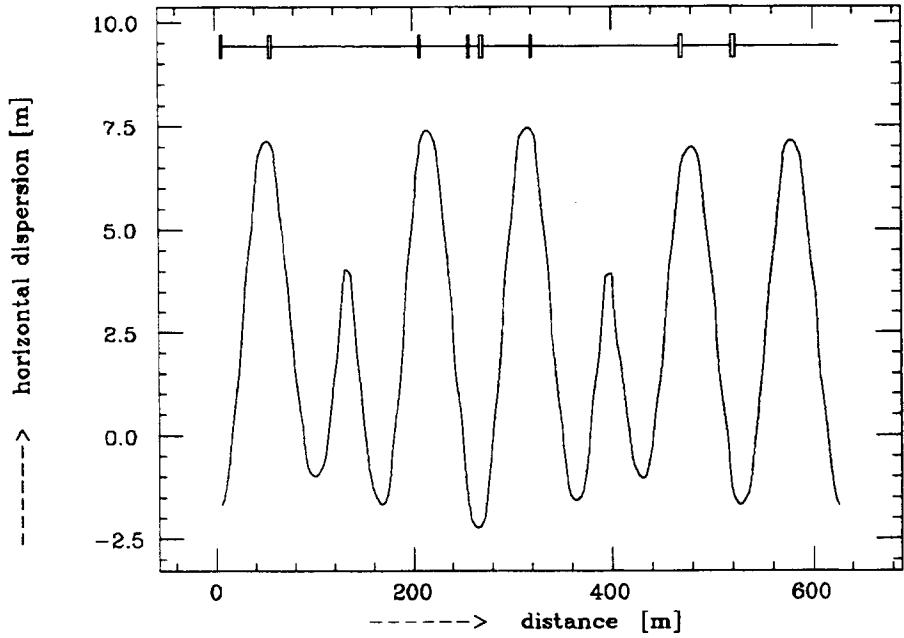


Figure 6: Dispersion in the PS machine perturbed by family 2 only (sampled at locations with equal  $\beta$ ), and optimum locations for family 1

## 5 CONCLUSION

Two simple expressions, (11) and (15), have been derived relating  $\Delta(\gamma_t^{-2})$  to the integrated normalized gradients  $K_i$  of the quadrupole scheme, using only the nominal unperturbed  $\beta$  and dispersion values at the quadrupoles and their phase advances. At this stage no assumptions have been made about the layout of the quadrupole configuration, and the formalism may be used for the design of any  $\gamma_t$  modification scheme. Pulsed jump schemes have to meet the additional requirement of zero tune shifts, preferably obtained by the use of doublets, each consisting of two quadrupoles with opposite polarity at  $\pi$  (or  $3\pi, 5\pi, \dots$ ) phase advance and  $\beta_1 K_1 + \beta_2 K_2 = 0$  in both transverse planes. In this case terms of order higher than 2 in  $K$  cancel in the expression for  $\Delta(\gamma_t^{-2})$ .

In machines where the horizontal normalized dispersion  $D/\sqrt{\beta}$  has the same value at all  $\gamma_t$  modification quadrupoles, the linear terms in  $K$  cancel for zero tune shifts. The remaining quadratic terms may easily be computed to allow the performance evaluation of the different design options, like the number of families and the superperiodicity. A bipolar  $\gamma_t$  jump provides a large jump amplitude while keeping the maximum  $\beta$  and  $D$  excursions from the nominal values at a reasonable level. This requires two quadrupole families, as is shown in the example of the CERN PS scheme. The distribution of the doublets around the ring, and more particularly the superperiodicity, has to be chosen with care in order to maximize the mixing term  $bK_1K_2$  in equation (24) which produces the bipolar jump.

On the other hand, in machines where the  $D/\sqrt{\beta}$  modulation is important a scheme may be found where  $\Delta(\gamma_t^{-2})$  has non-zero linear terms in  $K$  even for zero tune shifts, in which case a single quadrupole family is sufficient to produce a bipolar jump as is shown

in the example of the CERN SPS.

The difference between these two cases may more easily be understood by the following. Any machine with a *matched*  $D/\sqrt{\beta}$  function may be converted into a *mismatched* machine by powering one quadrupole family. A second family will then produce a  $\Delta(\gamma_t^{-2})$  which is linear in its strength value. Figure 6 shows that, as may be expected from the SPS case, in the PS the optimum locations for family 1 correspond to the minima and maxima of the dispersion in the presence of family 2.

The superperiodicity plays an important role in matched  $D/\sqrt{\beta}$  lattices. A machine may be divided into  $n$  identical superperiods, each having a tune  $Q/n$  which determines the  $m_{ij}$  values in machines where the quadratic terms in  $K$  dominate  $\Delta(\gamma_t^{-2})$ . Transparent betatron wavelengths without jump quadrupoles may be inserted between these superperiods, thus offering an additional choice of superperiod tunes like  $(Q - 1)/n$ ,  $(Q - 2)/n$  etc., as shown in the example of the CERN PS. To increase the efficiency of a scheme  $n$  may be chosen such that the superperiod tune is close to an integer in the quadratic case. Unfortunately  $D_{max}^*$  will generally also be larger in this case. On the other hand, machines where  $D/\sqrt{\beta}$  is not matched are dominated by the first term of (21) and do not show this resonant behaviour of  $\Delta(\gamma_t^{-2})$ .

## ACKNOWLEDGEMENT

The author gratefully acknowledges the support of W. Hardt during the preparation of this lecture.

## References

- [1] K. Johnsen, Transition, Proc. CERN Accelerator School 1984, Gif-sur-Yvette, CERN 85-19, November 1985
- [2] J. Le Duff, Longitudinal Beam Dynamics in Circular Accelerators, These proceedings
- [3] A. Sørenssen, Longitudinal Space-charge Forces at Transition, Proc. 6th Int. Conf. on High Energy Acc., Cambridge, Mass., 1967, (Cambridge Electron Accelerator, Cambridge, 1967).
- [4] A. Sørenssen, Crossing the Phase Transition in Strong Focussing Proton Synchrotrons, CERN MPS/DL/73-9, January 1974, also published in Particle Accelerators, Vol. 6, No. 3 (1973)
- [5] K. Johnsen, Proc. CERN Symposium on High Energy Acc. and Pion Physics, CERN, 1956, (CERN, Geneva, 1956) Vol. 5, p. 106
- [6] W. Hardt, Gamma Transition Jump Scheme of the CPS, Proc. 9th Int. Conf. on High Energy Acc., Stanford, April 1974, (CONF 740522, USAEC, Washington, 1974)
- [7] W.W. Lee and L.C. Teng, Proc. 8th Int. Conf. on High Energy Acc., Geneva, 1971 (CERN, Geneva, 1971)
- [8] D. Trbojevic, Design Method of High Energy Accelerator without Transition Energy, Proc. European Particle Acc. Conf., Nice, 1990 (Editions Frontières), Vol. 2, p. 1536
- [9] D. Möhl, Space Charge Effects at Transition Energy, 1973 PEP Summer Study, LBL/SLAC
- [10] J. Rossbach and P. Schmüser, Basic Course on Accelerator Optics, These proceedings
- [11] W. Hardt and D. Möhl, Q-jump at Transition, CERN ISR-300/GS/69-16, 1969
- [12] D. Möhl, Compensation of Space Charge Effects at Transition by an Asymmetric Q-jump - A Theoretical Study, CERN ISR-300/GS/69-62, 1969
- [13] L.C. Teng, Compensation of Space-Charge Mismatch at Transition of Booster using the Transition Jump Method, FN-207, 17 April 1970



# TUNE SHIFTS FROM SELF-FIELDS AND IMAGES

*A. Hofmann*

CERN, Geneva, Switzerland

## Abstract

The self-fields created by the particles in the beam are modified by the boundary conditions imposed by the beam surroundings. They act back on the beam and influence its motion and change the betatron frequencies. Part of this effect is caused by the direct Coulomb force between the particles. For a moving beam this Coulomb field is accompanied by a magnetic field. Since the forces due to these two fields have opposite signs there is a partial cancellation which becomes more perfect as the particle speed approaches that of light. This leads to the  $1/\gamma^2$  dependence of the direct space-charge force. For a smooth and perfectly conducting vacuum chamber the boundary conditions for the field can be satisfied by introducing image charges or currents. Their fields act on the beam and create an indirect space-charge effect. Since the electric and magnetic fields have quite different boundary conditions the cancellation between the electric and magnetic forces is disturbed. The indirect space-charge force does therefore not decrease with  $1/\gamma^2$  and can be dominant at high energy. If the beam as a whole does not oscillate the space-charge forces are stationary and affect only the focusing of the individual particles leading to the incoherent tune shifts. For a beam executing a center-of-mass motion the source of the space-charge fields changes resulting in a coherent tune shift which can be different from the incoherent one. This coherent effect is only important for the indirect space-charge fields since the direct effect is caused by internal forces which move with the beam. For a realistic beam the space-charge tune shift of an individual particle depends on its betatron amplitude and its longitudinal position along the bunch. As a consequence the space-charge forces lead to a tune spread. The beam occupies therefore a sizable area in the tune diagram leaving a limited space between the resonances.

## 1. INTRODUCTION

The motion of a single particle in an accelerator is determined by the external guide fields, i.e. the dipole and quadrupole magnets, the RF system, etc. A beam containing many particles represents a sizable charge and current which create fields of their own. They are modified by the boundary conditions imposed by the beam surroundings represented by the vacuum chamber, cavities etc. and act back on the beam. This influences the motion of the individual particles (incoherent effect) as well as of the beam as a whole (coherent effect).

We concentrate here on effects which influence the transverse particle motion and change the betatron frequencies. We also restrict ourself to simple cases which can be calculated without problems. More detailed investigations and more complicated cases can be found in some review papers [1, 2, 3]. Furthermore we assume here that the space-charge effect is weak and can be treated as a perturbation. Other methods must be used to describe space-charge dominated beams [4].

## 2. DIRECT INCOHERENT TUNE SHIFTS

### 2.1 Fields of a coasting beam with circular cross section

We consider now a coasting, i.e. unbunched beam with a circular cross section of radius  $a$  and uniform charge density  $\eta$  moving with constant velocity  $v = \beta c$  in the  $s$ -direction, Fig. 1. This beam represents a charge per unit length  $\lambda = \pi a^2 \eta$ , a current density  $J \beta c \eta$  and a total current  $I = \beta c \lambda = \pi a^2 J = \pi a^2 \beta c \eta$ . The charge produces an electric field  $\mathbf{E}$  and the current a magnetic field  $\mathbf{B}$  which are determined by the following relations:

$$\operatorname{div} \mathbf{E} = \eta / \epsilon_0$$

$$|\mathbf{E}| = E_r$$

$$\int \int \int \operatorname{div} \mathbf{E} dV = \int \int \mathbf{E} d\mathbf{S} \text{ (Gauss)}$$

$$\operatorname{curl} \mathbf{B} = \mu_0 \mathbf{J}$$

$$|\mathbf{B}| = B_\phi = \mu_0 \eta \beta c$$

$$\oint \mathbf{B} \cdot d\mathbf{s} = \int \int \operatorname{curl} \mathbf{B} \cdot d\mathbf{S} \text{ (Stokes).}$$

Here  $dV$  is the volume,  $d\mathbf{S}$  the surface and  $d\mathbf{s}$  the line element of the integration. We consider a cylinder of radius  $r$  and length  $\ell$  as shown in Fig. 1 and calculate the fields inside ( $r \leq a$ ) and outside ( $r \geq a$ ) the beam.

$$r \geq a$$

$$\pi \ell r^2 \eta / \epsilon_0 = 2\pi \ell r E_r$$

$$E_r = \frac{\eta}{2\epsilon_0} r = \frac{\lambda}{2\pi\epsilon_0 a^2} \frac{r}{a^2} = \frac{I}{2\pi\epsilon_0 \beta c} \frac{r}{a^2}$$

$$2\pi r B_\phi = \mu_0 \pi r^2 J_z = \pi \mu_0 \beta \eta c$$

$$B_\phi = \frac{\beta \eta}{2\epsilon_0 c} r = \frac{I}{2\pi\epsilon_0 c^2} \frac{r}{a^2}$$

$$r \leq a$$

$$\pi \ell a^2 \eta / \epsilon_0 = 2\pi \ell r E_r$$

$$E_r = \frac{\eta a^2}{2\epsilon_0 r} = \frac{\lambda}{2\pi\epsilon_0 r} \frac{1}{a^2} = \frac{I}{2\pi\epsilon_0 \beta c r} \frac{1}{a^2}$$

$$2\pi r B_\phi = \mu_0 \pi r^2 J_z = \mu_0 c \pi \beta \eta$$

$$B_\phi = \frac{\beta \eta}{2\epsilon_0 c} \frac{a^2}{r} = \frac{I}{2\pi\epsilon_0 c^2} \frac{1}{r}$$

We made the replacement  $\mu_0 = 1/(\epsilon_0 c^2)$  and used the formal relation  $\lambda = I/\beta c$  to express the line charge density  $\lambda$  by the current  $I$  which can be more easily measured. One should however remember that the electric field is created by charge and not by the current.

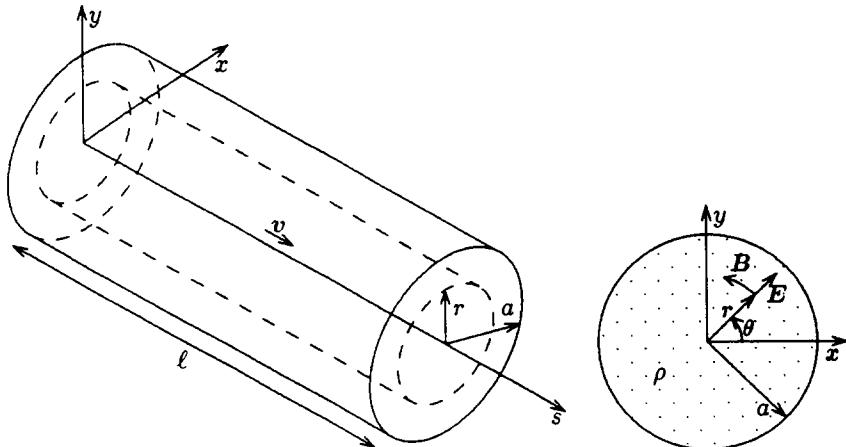


Figure 1: Space-charge fields in a uniform and circular beam

We calculate now the direct space-charge effect which is caused by the internal forces in the beam and apply the obtained results to calculate the space-charge forces on a particle with charge  $e$  inside the beam. From the fields

$$E_r = \frac{\eta}{2\epsilon_0}r, \quad B_\phi = \frac{\beta}{c}\frac{\eta}{2\epsilon_c}r$$

we get the force

$$\mathbf{F} = e(\mathbf{E} + [\mathbf{v} \times \mathbf{B}]) = \frac{e\eta}{2\epsilon_0} (1 - \beta^2) \mathbf{r} = \frac{e\eta}{2\epsilon_0 \gamma^2} \mathbf{r} = \frac{eI}{2\pi\epsilon_0 b\gamma^2} \frac{\mathbf{r}}{a^2}$$

where we introduced the Lorentz factor

$$\gamma = \frac{1}{\sqrt{1 - \beta^2}}.$$

The force is transverse and has the components

$$F_x = \frac{e\eta}{2\epsilon_0 \gamma^2} x, \quad F_y = \frac{e\eta}{2\epsilon_0 \gamma^2} y.$$

The electrostatic and the magnetic forces have opposite directions which leads to the partial cancellation between the two forces which becomes more perfect as the velocity of the particle approaches the one of light. This is one of the most important properties of the space-charge effects. For relativistic particles with  $\gamma \gg 1$  the direct space-charge force is relatively small. However, we will later investigate several effects which disturb the cancellation between the two forces.

## 2.2 The incoherent betatron tune shift due to the direct space-charge force

### 2.2.1 Coasting beam uniform density and focusing

The cross section of a beam circulating in a realistic machine has a different cross section in the focusing and defocusing quadrupoles. However, to start with the most

simple case, we approximate for a uniform focusing where the beta functions are constant and given by the betatron tunes  $Q_x$  and  $Q_y$ , i.e. the number of betatron oscillations per turn, and the average radius  $R$  of the ring

$$\beta_x \approx \langle \beta_x \rangle \approx \frac{R}{Q_x}, \quad \beta_y \approx \langle \beta_y \rangle \approx \frac{R}{Q_y}.$$

Furthermore we assume the beam to have a circular cross section of radius  $a$ , a uniform density and to be unbunched.

The motion of the particle in the beam in the absence of space-charge is determined by the linear focusing provided by the quadrupoles and is described by the equation

$$\frac{d^2y}{dt^2} + \omega_\beta^2 y = \ddot{y} + Q_{y0}^2 \omega_0^2 y = 0,$$

where  $\omega_\beta$  is the betatron frequency and  $\omega_0 = \beta c/R$  the revolution frequency. The solution is a betatron oscillation of the form

$$y(t) = y_0 \cos(Q_{y0}\omega_0 t).$$

The presence of the space-charge force imposes an additional acceleration to the individual particles in the beam

$$\ddot{y}_{s.c.} = \frac{F_y}{m_0\gamma} = \frac{e\eta}{2\epsilon_0\gamma^2 m_0\gamma} y$$

which has to be included and leads to the equation of motion

$$\ddot{y} + Q_{y0}^2 \omega_0^2 y = \frac{e\eta}{2\epsilon_0 m_0 \gamma^3} y,$$

or

$$\ddot{y} + \left( Q_{y0}^2 \omega_0^2 - \frac{e\eta}{2\epsilon_0 m_0 \gamma^3} \right) y = 0.$$

The term in the large brackets represents the square of the new betatron frequency which includes the space-charge force. We call the corresponding tune  $Q_y$  which is related to the unperturbed tune  $Q_{y0}$  by

$$Q_y^2 = Q_{y0}^2 \left( 1 - \frac{e\eta}{2\epsilon_0 Q_{y0}^2 \omega_0^2 m_0 \gamma^3} \right) y. = Q_{y0}^2 \left( 1 - \frac{eI}{2\pi\epsilon_0 Q_{y0}^2 \omega_0^2 a^2 m_0 \beta c \gamma^3} \right) y.$$

Since the space-charge force is linear in  $y$  like the external focusing force it just leads to a change  $\Delta Q_y$  of the betatron tune of the individual particle. We assume now that this change is small,  $\Delta Q_y \ll Q_{y0}$ , such that the quadratic term in it can be neglected

$$Q_y^2 = (Q_{y0} + \Delta Q)^2 \approx Q_{y0}^2 + 2Q_{y0}\Delta Q$$

and get for the incoherent tune shift

$$\Delta Q_y \approx -\frac{e\eta}{4\epsilon_0 Q_{y0} \omega_0^2 m_0 \gamma^3} = -\frac{ecI}{4\pi\epsilon_0 Q_{y0} \omega_0^2 a^2 m_0 c^2 \beta \gamma^3}$$

We introduce the classical electron radius

$$r_0 = \frac{e^2}{4\pi\epsilon_0 m_0 c^2} = \begin{cases} 1.54 \cdot 10^{-18} \text{ m for protons} \\ 2.82 \cdot 10^{-15} \text{ m for electrons} \end{cases}$$

and get

$$\Delta Q_y = -\frac{r_0 c I}{e \omega_0^2 Q_{y0} a^2 \beta \gamma^3}.$$

We approximated for uniform focusing with a beta function  $\beta_y$  and an emittance  $E_y$

$$\beta_y \approx \langle \beta_y \rangle \approx \frac{R}{Q_{y0}}, \quad E_y \approx \frac{a^2}{\beta_y} \approx \frac{Q_{y0}}{R} a^2.$$

With this we get an expression for the incoherent tune shift which uses more practical machine and beam parameters

$$\Delta Q_y = -\frac{r_0 R I}{e c E_y \beta^3 \gamma^3}, \quad \Delta Q_x = -\frac{r_0 R I}{e c E_x \beta^3 \gamma^3},$$

where we also give the corresponding expression for the horizontal plane.

In deriving these equation we made several assumptions like uniform focusing, circular beams, unbunched beams and uniform density. We will now have a look at these restrictions.

### 2.2.2 Non-uniform focusing

We can derive the space-charge tune shift differently by applying an expression commonly used in beam optics for the tune shift due to a localized weak lens

$$\Delta Q_x = \frac{K_x \ell \beta_y}{4\pi},$$

where  $\ell$  is the length of the lens and  $K_x$  the focusing parameter which is given by

$$K_x = \frac{e}{m_0 c \beta \gamma} \frac{dB_y}{dx}$$

for a magnetic quadrupole lens or by

$$K_x = -\frac{d(1/\rho)}{dx} \approx -\frac{1}{x} \frac{d^2 x}{ds^2}$$

for the general case with the curvature  $1/\rho \approx d^2 x/ds^2$  of the trajectories in a linear lens. The space-charge force gives an acceleration

$$\ddot{x} = \frac{F_x}{m_0 \gamma} \approx \frac{d^2 x}{ds^2} \beta^2 c^2 = \frac{2 r_0 c I}{e a^2 \beta \gamma^3} x$$

which gives for the focusing parameter

$$K_x = \frac{2 r_0 I}{\pi c a^2 \beta^3 \gamma^3}.$$

We regard now a beam disk of length  $ds$  as a short lens which produces a tune change

$$d(\Delta Q_x) = \frac{\beta_x K_x}{4\pi} ds = -\frac{r_0 I \beta_x}{2 \pi e c a^2 \beta^3 \gamma^3} ds = -\frac{r_0 I}{2 \pi e c E_x \beta^3 \gamma^3} ds.$$

The emittance  $E_x = a^2/\beta_x$  is invariant and we integrate around the ring and get for the total space-charge tune shift the same expression as before

$$\Delta Q_x = - \int_0^{2\pi R} \frac{r_0 I}{2\pi e c E_x \beta^3 \gamma^3} ds = -\frac{r_0 I R}{e c E_x \beta^3 \gamma^3}.$$

The corresponding expression for the horizontal plane is of the same form since we assumed circular beams.

It is interesting to note that the direct space-charge tune shift depends on the emittance and not on the  $\beta$ -function (local focusing) or the tune  $Q_x$  (global focusing). At places where the beta function is small the beam is small too which results in a strong space-charge force. On the other hand a given lens produces a smaller tune change if it is located at a small value of the beta function. However, we assumed here a beam being circular all around the ring which is not realistic and will be discussed later. The tune shift decreases with the third power of the Lorentz factor  $\gamma$ . The cancellation between the magnetic and electric forces contributes a power of two and the stiffness of the beam a power of one to this dependence. For a given current the space-charge effect increases with the radius  $R$  of the machine. However, we can eliminate this dependence by expressing the current with the number of circulating particles  $N_b$

$$I = \frac{e N_b \omega_0}{2\pi} = \frac{e N_b \beta c}{2\pi R}$$

which gives

$$\Delta Q_x = -\frac{r_0 N_b}{2\pi E_x \beta^2 \gamma^3} = -\frac{eq_b}{8\pi^2 \epsilon_0 E_x E \beta^2 \gamma^2},$$

where  $e$  is the charge of a particle and  $q_b = e N_b$  of the whole beam and  $E = m_0 c^2 \gamma$  the particle's energy.

### 2.2.3 Uniform elliptic beam

We generalize the above result for a beam with elliptic cross section with half-axes  $a$  and  $b$  but still with uniform charge density  $\eta$ . The electric field inside such a beam is given by [5]

$$\mathbf{E} = \frac{I}{\pi \epsilon_0 (a + b) \beta c} \left( \frac{x}{a}, \frac{y}{b}, 0 \right), \quad \mathbf{B} = \frac{\mu_0 I}{\pi (a + b)} \left( -\frac{y}{b}, \frac{x}{a}, 0 \right)$$

which can be checked to be consistent with  $\text{div } \mathbf{E} = \eta / \epsilon_0$  and  $\text{curl } \mathbf{B} = \mu_0 \mathbf{J}$ . The space-charge force becomes

$$\mathbf{F} = e (\mathbf{E} + [\mathbf{v} \times \mathbf{B}]) = \frac{I}{\pi \epsilon_0 \beta c \gamma^2 (a + b)} \left( \frac{x}{a}, \frac{y}{b}, 0 \right)$$

and the space-charge tune shifts

$$\begin{aligned}\Delta Q_x &= - \oint \frac{2r_0 I}{2\pi ec\beta^3\gamma^3} \frac{\beta_x}{(a+b)a} ds = - \frac{2r_0 IR}{ec\beta^3\gamma^3} \frac{1}{E_x} \left\langle \frac{a}{a+b} \right\rangle \\ &= - \frac{r_0 IR}{ec\beta^3\gamma^3 E_x} \left\langle \frac{2}{1 + \sqrt{\kappa\beta_y/\beta_x}} \right\rangle \\ \Delta Q_y &= - \oint \frac{2r_0 IR}{2\pi ec\beta^3\gamma^3} \frac{\beta_y}{(a+b)b} ds = - \frac{2r_0 IR}{ec\beta^3\gamma^3} \frac{1}{E_y} \left\langle \frac{b}{a+b} \right\rangle \\ &= - \frac{r_0 IR}{ec\beta^3\gamma^3 E_y} \left\langle \frac{2}{1 + \sqrt{\beta_x/(\kappa\beta_y)}} \right\rangle,\end{aligned}$$

where  $\kappa = E_y/E_x$  is the ratio between the vertical and horizontal emittances, also called coupling factor. We see that in the case of elliptic beams the space-charge tune shift depends somewhat on the beam optics of the ring.

#### 2.2.4 Bunched beams

For most applications we deal with bunched beams and the current  $I(s - s_0)$  is a function of the longitudinal position  $s$  from the bunch center  $s_0$ . This leads to some longitudinal force and can make the space-charge effect more complicated. However, for relativistic beams a simple approximation can be used. The electric field of a particle moving with velocity  $v = \beta c$  is mainly transverse with an opening angle of order  $1/\gamma$ . The transverse forces seen by a particle in the beam are mostly caused by the part of the beam which is within a longitudinal distance  $\Delta s \leq a/\gamma$ . If the bunch is sufficiently long that the current  $I(s)$  changes little over this distance  $\Delta s$  we get the direct space-charge tune shift of a bunched beam by replacing in the coasting beam expressions the beam current  $I$  by its local value  $I(s)$

$$\Delta Q_y(s - s_0) = - \frac{r_0 RI(s - s_0)}{ecE_y\beta^3\gamma^3}, \quad \Delta Q_x(s - s_0) = - \frac{r_0 RI(s - s_0)}{ecE_x\beta^3\gamma^3}.$$

The space-charge tune shift of a particle depends now on its longitudinal position in the bunch. This leads to a tune spread in the bunch and, since the particle executes synchrotron oscillations, to a modulation of the tune.

#### 2.2.5. Non-uniform charge density

So far we assumed that the charge density  $\eta$  in the beam is constant which gave a linear defocusing force. For a more general transverse charge distribution this force is no longer linear and the tune shift will depend on the betatron amplitude of a particle. We illustrate this for a circular beam with a Gaussian density distribution

$$\eta(r) = \frac{I}{2\pi\beta c\sigma^2} e^{-\frac{r^2}{2\sigma^2}}$$

which leads to the field components

$$E_r = \frac{I}{2\pi\epsilon_0\beta c r} \frac{1}{r} \left( 1 - e^{-\frac{r^2}{2\sigma^2}} \right), \quad B_\phi = \frac{I}{2\pi\epsilon_0 c^2 r} \frac{1}{r} \left( 1 - e^{-\frac{r^2}{2\sigma^2}} \right)$$

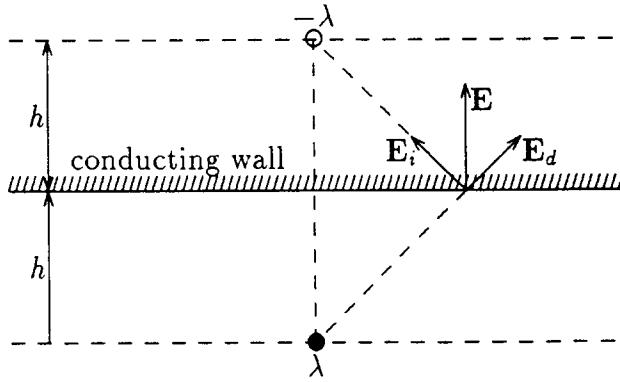


Figure 2: Satisfying an electric boundary condition by an image charge

and to the radial force

$$F_r = \frac{eI}{2\pi\epsilon_0\beta c\gamma^2} \frac{1}{r} \left( 1 - e^{-\frac{r^2}{2\sigma^2}} \right).$$

This space-charge force inside the Gaussian beam is non-linear and can be developed into powers of  $r$

$$F_r \approx \frac{eI}{2\pi\epsilon_0\beta c\gamma^2} r \left( 1 - 1 + \frac{r^2}{2\sigma^2} - \frac{1}{2} \frac{r^4}{4\sigma^4} + \dots \right) = \frac{eI}{2\pi\epsilon_0\beta c\gamma^2} \left( \frac{r^2}{2\sigma^2} - \frac{r^3}{8\sigma^4} + \dots \right).$$

Taking only the first term we get the small amplitude tune shift

$$\Delta Q_x \approx \frac{r_0 I R}{e \beta^3 \gamma^3} \frac{\beta_x}{2\sigma^2}$$

in a Gaussian beam which is the same as the one of an uniform beam with radius  $a = \sqrt{2}\sigma$ .

### 3. EFFECTS OF THE WALLS ON THE INCOHERENT TUNE SHIFT

#### 3.1 Conducting parallel plates

The presence of a perfectly conducting plate close to the beam imposes a boundary condition  $E_{||} = 0$  on its surface. This modifies the electric field of the beam. It is possible to solve the electrodynamic equations for the new boundary conditions. There is a relatively simple method to do this by introducing image charges. We consider now a beam which has at a distance  $h$  a horizontal infinite plate which is perfectly conducting. We introduce an image charge with the same magnitude of line density as the beam but of opposite sign, Fig. 2. On the surface of the plate the electric field  $\mathbf{E}_i$  of the image is such that it cancels the parallel field component of the beam itself. The total field in front of the plate consists of the direct part  $\mathbf{E}_d$  and the part created by the images

$$\mathbf{E} = \mathbf{E}_d + \mathbf{E}_i, \quad E_{||} = E_{d||} + E_{i||} = 0.$$

The image charge is all behind the plate and its divergence in front of the plate vanishes

$$\operatorname{div} \mathbf{E} = \operatorname{div} \mathbf{E}_d + \operatorname{div} \mathbf{E}_i = \operatorname{div} \mathbf{E}_d = \frac{\eta}{\epsilon_0}.$$

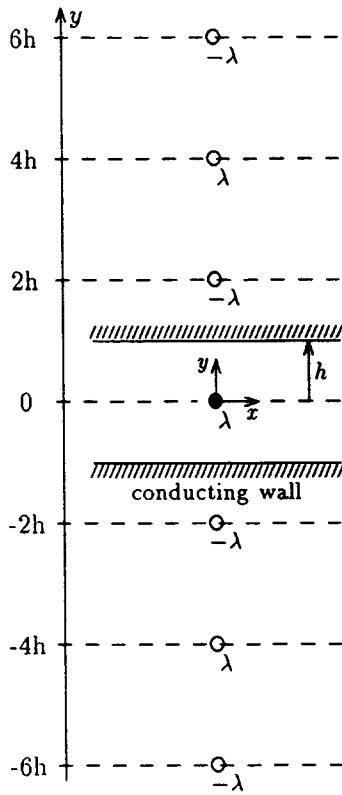


Figure 3: Incoherent tune shift in the presence of conducting plates

In front of the plate the total field fulfills therefore Maxwell's equations and the boundary condition.

We would now like to calculate the effect of the beam surroundings on the space-charge effect. We consider first a flat vacuum chamber and approximate it by two parallel conducting plates at the vertical distance  $\pm h$  from the beam. To fulfill the boundary conditions we introduce first a negative line charge at the distance  $h$  behind the upper and lower plate, i.e. at the distance  $\pm 2h$  from the beam. However, the field created by the image of the upper plate is also influenced by the boundary condition of the lower plate. This can be satisfied by introducing a positive image line charge at  $3h$  below the lower plate. To fulfill all these conditions we end up with negative images at the distances  $\pm 2nh$  with  $n = 1, 3, 5, \dots$  and positive images at  $\pm 2nh$  with  $n = 2, 4, 6, \dots$  as shown in Fig. 3.

We calculate the field of each image  $\lambda_{in}$  as a function of the coordinates  $x$  and  $y$  in the vicinity of the beam. Since the images are relatively far away from the beam compared to its size,  $h \gg a$ , we neglect quadratic and higher terms in  $x$  and  $y$ . The field of the  $n$ th image pair at the distance  $\pm 2nh$  is

$$E_{iny} = (-1)^n \frac{\lambda}{2\pi\epsilon_0} \left( \frac{1}{2nh + y} - \frac{1}{2nh - y} \right) = (-1)^n \frac{\lambda}{2\pi\epsilon_0} \frac{-2y}{(2nh)^2} = -\frac{\lambda y}{4\pi\epsilon_0 h^2} \frac{(-1)^n}{n^2}.$$

$$E_{iny} = -\frac{\lambda y}{4\pi\epsilon_0 h^2} \sum_1^{\infty} \frac{(-1)^n}{n^2} = \frac{\lambda y}{4\pi\epsilon_0 h^2} \frac{\pi^2}{12}.$$

To get the field of all charges we summed over  $n$  and used the expression

$$\sum_1^{\infty} \frac{(-1)^n}{n^2} = -\frac{\pi^2}{12}$$

which can be found in standard text books. The horizontal field could be evaluated the same way but we can get it directly from

$$\operatorname{div} \mathbf{E}_i = \frac{\partial E_{ix}}{\partial x} + \frac{\partial E_{iy}}{\partial y} = 0$$

from which one gets

$$\frac{\partial E_{ix}}{\partial x} = -\frac{\partial E_{iy}}{\partial y} = -\frac{\lambda}{4\pi\epsilon_0 h^2} \frac{\pi^2}{12},$$

or

$$E_{ix} = -\frac{\lambda y}{4\pi\epsilon_0 h^2} \frac{\pi^2}{12}.$$

The boundary condition imposed by the conducting plates does not affect the magnetic field. As a consequence there is no cancellation effect for the forces due to the images. The total force is therefore

$$\begin{aligned} F_x &= \frac{2e\lambda y}{2\pi\epsilon} \left( \frac{1}{2a^2\gamma^2} - \frac{\pi^2}{48h^2} \right) \\ F_y &= \frac{2e\lambda y}{2\pi\epsilon} \left( \frac{1}{2a^2\gamma^2} + \frac{\pi^2}{48h^2} \right) \end{aligned}$$

and the resulting tune shift

$$\begin{aligned} \Delta Q_x &= -\frac{2r_0IR\langle\beta_x\rangle}{ec\beta^3\gamma} \left( \frac{1}{2a^2\gamma^2} - \frac{\pi^2}{48h^2} \right) \\ \Delta Q_y &= -\frac{2r_0IR\langle\beta_y\rangle}{ec\beta^3\gamma} \left( \frac{1}{2a^2\gamma^2} + \frac{\pi^2}{48h^2} \right), \end{aligned}$$

where we used again the formal substitution  $I = \lambda\beta c$ .

The first term in the large brackets is the direct space-charge effect which decreases as  $1/\gamma^2$  due to the cancellation between the forces. It also depends on the beam size, i.e. on the charge density. The second term gives the indirect effect of the wall which does not have the cancellation and which does not depend on the beam size within our approximation  $h \gg a$  but on the distance  $h$  of the plate. In addition both effects decrease as  $1/\gamma$  due to the particles becoming more rigid at higher energy.

### 3.2 Ferromagnetic boundary

We look now at ferromagnetic boundaries which affect the magnetic field and can be treated in a way similar to the conducting plates. They appear mostly in the form of the dipole magnet pole pieces and can be approximated by a horizontal ferromagnetic sheet at a distance  $g$ . They impose the boundary condition  $B_{||} = 0$  for the magnetic field. This can be satisfied by a positive image current at a distance  $g$  behind the sheet, i.e. at  $2g$  from the beam; Fig. 4. The original and the image current produce fields on

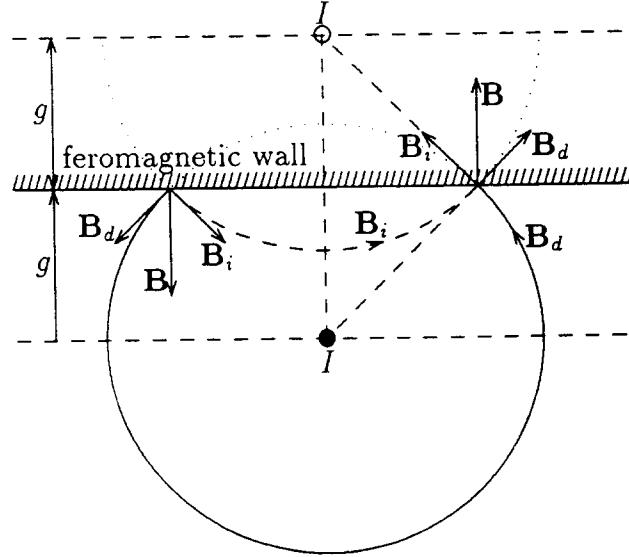


Figure 4: Satisfying a magnetic boundary condition by an image current

the ferromagnetic surface with opposite parallel components. To approximate a realistic magnet we need a sheet above and below the beam which makes it necessary to introduce image currents at  $\pm 2gh$  from the beam similarly to the conducting plates but here all image currents have the same sign as the beam itself. Using again only linear terms in  $x$  and  $y$  we get for the magnetic field produced by these image currents

$$B_{ix} = \frac{\mu_0 I}{2\pi} \sum_1^{\infty} \left( \frac{1}{2ng - y} - \frac{1}{2ng + y} \right) = \frac{\mu_0 I y}{4\pi g^2} \sum_1^{\infty} \frac{1}{n^2} = \frac{\mu_0 I y \pi^2}{4\pi g^2 6},$$

where we used the expression

$$\sum_1^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}.$$

Including this we get for the total vertical force

$$F_y = \frac{2eIy}{2\pi\epsilon\beta c} \left( \frac{1}{2a^2\gamma^2} + \frac{\pi^2}{48h^2} + \frac{\pi^2\beta^2}{24g^2} \right)$$

and for the tune shift

$$\Delta Q_y = -\frac{2r_0IR\langle\beta_y\rangle}{\epsilon c\beta^3\gamma} \left( \frac{1}{2a^2\gamma^2} + \frac{\pi^2}{48h^2} + \frac{\pi^2\beta^2}{24g^2} \right).$$

To get the horizontal tune shift due to this horizontal ferromagnetic sheet we use the same reasoning as before

$$\operatorname{div} \mathbf{B} = \frac{\partial B_y}{\partial x} + \frac{\partial B_x}{\partial y} = 0$$

from which we get the vertical magnetic field

$$B_{iy} = -\frac{\mu_0 I x \pi^2}{4\pi g^2 6},$$

and the tune shift

$$\Delta Q_x = -\frac{2r_0IR(\beta_x)}{ec\beta^3\gamma} \left( \frac{1}{2a^2\gamma^2} - \frac{\pi^2}{48h^2} - \frac{\pi^2\beta^2}{24g^2} \right).$$

The ferromagnetic sheet influences the magnetic field. However, realistic pole pieces are usually also conducting and can affect the electric field also. In most practical cases the conducting vacuum chamber is between the beam and the pole piece ( $g > h$ ) and not penetrated by the electric field.

## 4. COHERENT TUNE SHIFTS

### 4.1 Coherent transverse oscillations

So far we treated the effect of the space-charge on the incoherent oscillation of individual particles inside the beam. In this case the center of mass is at rest, i.e. the source of the space-charge fields does not move. Now we investigate the effect on coherent oscillations in which the beam as a whole executes a betatron oscillation such that all particles move together in phase. The space-charge fields are now modulated with this motion and act back on the beam with a well determined phase. For this reason the coherent tune shift can be quite different from the incoherent one.

### 4.2 Coherent tune shift in free space

In the absence of any walls there are only the direct space-charge forces acting on the beam. If the beam executes a coherent oscillation these space-charge forces are internal and move with the beam. They do not affect the motion of the beam as a whole and the coherent tune is not changed. However, the internal motion of the individual particles with respect to the center of mass is influenced by the direct space-charge force in the same way as for a beam at rest. Sometimes one calls the coherent tune shift the difference between the coherent and the incoherent tune. Since the coherent tune does not change we find for the free space case

$$Q_{coh.} = Q_0 , \quad Q_{incoh.} = Q_0 + \Delta Q , \quad \delta Q_{coh.} = Q_{coh.} - Q_{inc.} = -\Delta Q.$$

### 4.3 Coherent tune shifts for conducting parallel plates

We come back to the approximation of the vacuum chamber by conducting plates treated before but we assume now that the beam makes a coherent motion and has at present a vertical deviation  $\bar{y}$ . Since the beam as a whole is displaced the images move also. We calculate the field produced by the first image pair at the position  $y = \bar{y}$  of the beam. The upper image is at the nominal distance  $2h$ . If the beam is displaced upwards by  $\bar{y}$  the image moves towards the beam by the same amount such that the distance from the beam to the upper image becomes  $2h - 2\bar{y}$ , and to the lower image  $2h + 2\bar{y}$  giving a field at the beam, Fig. 5,

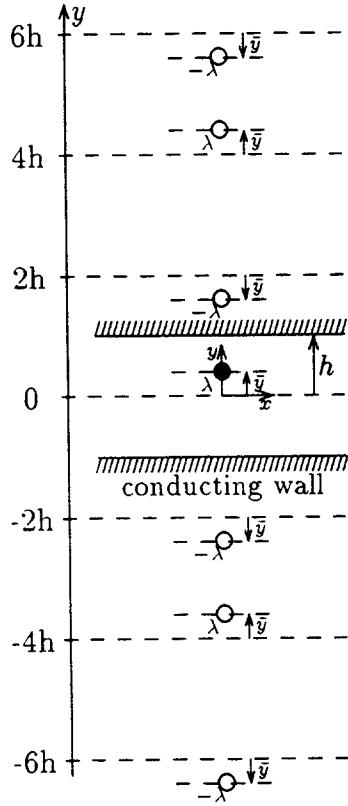


Figure 5: Coherent tune shift in the presence of conducting plates

$$E_{c1y} = \frac{-\lambda}{2\pi\epsilon_0} \left( \frac{1}{2h + \bar{y} + \bar{y}} - \frac{1}{2h - \bar{y} - \bar{y}} \right).$$

The second upper-image charge, which is really an image of the first lower-image charge, moves in the same direction as the beam such that its distance from the beam does not change. The field due to the second image pair can be expressed as

$$E_{c2y} = \frac{\lambda}{2\pi\epsilon_0} \left( \frac{1}{2h + \bar{y} - \bar{y}} - \frac{1}{2h - \bar{y} + \bar{y}} \right).$$

Generalizing this we get the field of the n-th image pair at the location of the displaced beam

$$\begin{aligned} E_{cn_y} &= \frac{(-1)^n \lambda}{2\pi\epsilon_0} \left( \frac{1}{2h + \bar{y}(1 - (-1)^n)} - \frac{1}{2h - \bar{y}(1 - (-1)^n)} \right) \\ &= -\frac{(-1)^n \lambda \bar{y}}{4\pi\epsilon_0 h^2} \left( \frac{1}{n^2} - \frac{(-1)^n}{n^2} \right). \end{aligned}$$

The field of all image charges is

$$E_{cn_y} = \frac{\lambda}{2\pi\epsilon_0 h^2} \left( -\sum_1^\infty \frac{(-1)^n}{n^2} + \sum_1^\infty \frac{1}{n^2} \right) = \frac{\lambda}{2\pi\epsilon_0 h^2} \left( \frac{\pi^2}{12} + \frac{\pi^2}{6} \right),$$

and the coherent force

$$F_{cy} = \frac{e\lambda\bar{y}}{\pi\epsilon_0 h^2} \frac{\pi^2}{16}.$$

This gives for the coherent tune

$$Q_{ycoh.} = Q_0 - \frac{\pi^2}{16} \frac{2r_0 IR(\beta_y)}{ec\beta^3\gamma h^2}.$$

The difference between coherent and incoherent tune becomes

$$\delta Q_{ycoh.} = Q_{ycoh.} - Q_{yinc.} = \frac{\pi^2}{16} \frac{2r_0 IR(\beta_y)}{ec\beta^3\gamma} \left( \frac{1}{2a^2\gamma^2} \frac{\pi^2}{48h^2} - \frac{\pi^2}{16h^2} \right).$$

For the horizontal conducting plates considered here a coherent horizontal motion is not influenced and there is no coherent horizontal tune change  $Q_{xcoh.} = Q_0$ . For a more realistic boundary condition like the one imposed by an elliptic chamber this is no longer the case.

#### 4.4 Circular chamber

We consider now a beam surroundend by a conducting circular cylinder of radius  $\rho$ . For the equilibrium case the beam is assumed to go along the  $z$ -axis. As a result a uniform surface charge of  $-\lambda$  is induced on the inside of this cylinder with a total charge per unit length. The field of this symmetrically distributed charge as well as its derivative vanish on the axis. Therefore the presence of the conducting cylinder does not affect the incoherent space-charge tune shift of the centered beam.

Next we assume the beam to be displaced in the horizontal direction by  $\bar{x}$ . The induced surface charges are no longer uniformly distributed and will produce a field at the beam. We try to satisfy the boundary condition on the surface by introducing an image line charge at the distance  $b$  from the axis. For symmetry reason it must lie in the  $x - z$  plane as the displaced beam. Using cylindrical coordinates  $\rho, \phi, z$  we get for the boundary condition

$$E_{||} = E_\phi = 0.$$

According to the upper part of Fig. 6 the beam produces such a field component

$$E_{\phi d} = \frac{\lambda}{2\pi\epsilon_0 r_1} \sin \alpha_1,$$

with

$$r_1^2 = \rho^2 + \bar{x}^2 - 2\rho\bar{x}\cos\phi, \quad \frac{\sin \alpha_1}{\sin \phi} = \frac{\bar{x}}{r_1}.$$

we get

$$E_{\phi 1} = \frac{\lambda}{2\pi\epsilon_0 \rho^2} \frac{\bar{x}\sin\phi}{\rho^2 + \bar{x}^2 - 2\rho\bar{x}\cos\phi}.$$

The field due to the image is obtained with help of the lower part of Fig. 7.

$$E_{\phi i} = -\frac{\lambda}{2\pi\epsilon_0 r_2} \sin \alpha_2,$$

with

$$r_2^2 = \rho^2 + b^2 - 2\rho b\cos\phi, \quad \frac{\sin \alpha_2}{\sin \phi} = \frac{b}{r_2}$$

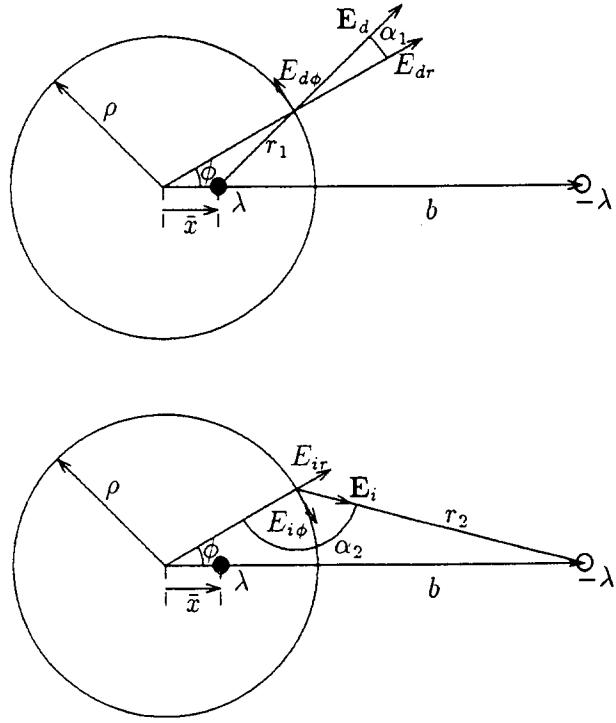


Figure 6: Coherent tune shift in a circular chamber

we get

$$E_{\phi 2} = -\frac{\lambda}{2\pi\epsilon_0} \frac{b \sin \phi}{\rho^2 + b^2 - 2\rho b \cos \phi}.$$

The condition  $E_{\phi 1} + E_{\phi 2} = 0$  leads to

$$\frac{\bar{x} \sin \phi}{\rho^2 + \bar{x}^2 - 2\rho \bar{x} \cos \phi} = \frac{b \sin \phi}{\rho^2 + b^2 - 2\rho b \cos \phi} =$$

or

$$\bar{x}b^2 - b(\bar{x}^2 + \rho^2) + \bar{x}\rho^2 = 0.$$

This equation does not depend on the angle  $\phi$  anymore which means that we can satisfy the boundary conditions with a single line charge at a distance  $b$ . We get for the latter

$$b_1 = \frac{\rho^2}{\bar{x}} , \quad b_2 = \bar{x}.$$

The image field at the beam has only an  $x$  component

$$E_{ix}(\bar{x}) = \frac{\lambda}{2\pi\epsilon_0} \frac{\bar{x}}{\rho^2}$$

which leads to a coherent tune shift

$$Q_{coh.} - Q_0 = \frac{1}{2} \frac{2r_0IR^2\langle\beta_x\rangle}{ec\beta^3\gamma\rho^2}.$$

Due to circular symmetry one gets the same tune shift in the vertical plane.

#### 4.5 The Laslett coefficients

We have calculated the indirect space-charge effect for the cases of a conducting and a ferromagnetic plate only. For more general beam surroundings the expressions have a similar form but with different numerical factors. We can write the incoherent tune shift in the form

$$\Delta Q_y = -\frac{2r_0IR\langle\beta_y\rangle}{ec\beta^3\gamma} \left( \frac{1}{2a^2\gamma^2} + \epsilon_1^V \frac{1}{h^2} + \epsilon_2^V \frac{1}{g^2} \right)$$

where  $\epsilon_1^V$  and  $\epsilon_2^V$  are the electric and magnetic Laslett space-charge coefficients for the vertical plane. The coherent shift can be written in a similar form with the coefficients  $\xi_1^V$  and  $\xi_1^H$ .

We get for the conducting plates at a distance  $h$

$$\begin{aligned}\epsilon_1^V &= \frac{\pi^2}{48}, \quad \epsilon_1^H = -\frac{\pi^2}{48} \\ \epsilon_2^V &= \frac{\pi^2}{24}, \quad \epsilon_2^H = -\frac{\pi^2}{24} \\ \xi_1^V &= \frac{\pi^2}{16}, \quad \xi_2^H = 0\end{aligned}$$

and for a circular chamber of radius  $h$ :

$$\begin{aligned}\epsilon_1^V &= \epsilon_1^H = 0 \\ \xi_1^V &= \xi_2^H = 0.5.\end{aligned}$$

### 5. COMMENTS

#### 5.1 Incoherent tune shifts

- The direct space-charge changes the incoherent betatron tunes. The cancellation between the electric and magnetic forces is very important for this effect and results in it decreasing like  $1/\gamma^2$  with energy. Another factor of  $1/\gamma$  is due to the relativistic mass increase. For unbunched beams with a circular cross section of radius  $a$  and uniform density we find

$$\Delta Q = -\frac{r_0IR\langle\beta_y\rangle}{ec\beta^3\gamma^3a^2}.$$

This result is obtained by assuming the space-charge defocusing is weak and can be treated as a first-order perturbation. For the case of beams which are dominated by space-charge other methods have to be used.

- For bunched beam the space-charge force depends on the longitudinal distance from the bunch center. This leads to a modulation of the tune for a particle executing a synchrotron oscillation and to a tune spread. In a beam with a non-uniform density the space-charge tune shift depends on the amplitude of the betatron oscillation.
- The beam surroundings impose some boundary conditions on the space-charge fields. Since they are in general different for the electric and magnetic fields the cancellation between the two forces is perturbed. This can lead to important tune shifts even at high energies.
- Beams can get partially neutralized by ions (electron and antiproton beams) or by electrons (proton beams). In this case the electric space-charge field is reduced and the cancellation between the forces perturbed leading to important tune shifts even for relativistic electron beams.
- In colliding beams the electric and magnetic forces produced by the other beam have the same direction and do not cancel. This leads to the well known beam-beam tune shift which has otherwise properties similar to the direct space-charge tune shift in a single beam.

## 5.2 Coherent effects

A betatron oscillation of the beam as a whole represents a coherent motion. This has little effect on the direct space charge effects which are determined by the internal forces inside the beam. However, the image charges and currents induced in the wall are modulated by the coherent beam oscillation. This leads to coherent tune shifts which can be quite different from the incoherent ones.

## 5.3 Consequences and problems caused by space-charge tune shifts

In the presence of optical machine imperfections the values of the betatron tunes should not be on or close to a rational number of low order otherwise some resonances are excited. The space-charge force can produce a large tune spread and it is more difficult to find a place in the tune diagram sufficiently far from the resonance.

The separation between the coherent and incoherent betatron frequency reduces the space available in the tune diagram where neither of the two tunes is close to a resonance. Furthermore the separation between the tunes can reduce the beneficial effect of Landau damping which is usually present for beams with a frequency spread.

## 5.4 Illustration of a difference between coherent and incoherent frequencies

The space-charge force produces a condition in which the coherent and incoherent betatron frequencies are different. Since this is a situation which is difficult to visualize we try to illustrate it by a simple example. We consider a swing with several children playing as shown in Fig. 7. If the oscillations of the individual swings are arbitrary they represent an incoherent motion and the frame of the swing does not move; top picture. The oscillation frequency is then given by  $\omega_{so} = \sqrt{g/L}$ , where  $g$  is the gravitational acceleration and  $L$  is the length of the swing.

If now all the children oscillate their swings in phase we have a coherent oscillation which applies a force on the frame. If this frame is not rigid it will move slightly with the coherent oscillation and change the frequency. This example is different from the cases of space-charge in so far as there are no direct forces between the individual swings. They are only coupled together via the frame which takes the role of the wall in the space-charge case.

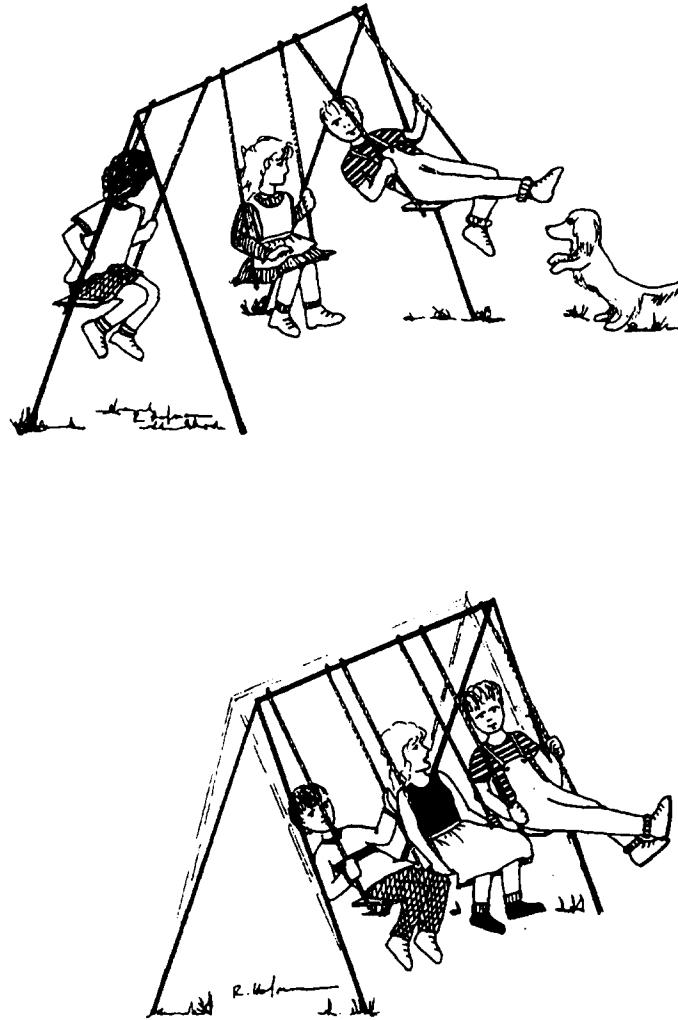


Figure 7: Example of a situation with different coherent and incoherent frequencies<sup>1</sup>

## 6. PRACTICAL EXAMPLES

We have a brief look at two practical examples of space-charge effects in accelerators and their consequences.

First the direct space-charge effect in the PS-booster [6]. This machine is a synchrotron which accelerates proton bunches from 50 to 800 MeV in about 0.6 s. The tunes occupied by the particles are indicated in the diagram of Fig. 8 by the shaded areas. At injection ( $t = 0$  ms) the space-charge effect is strongest as indicated by the lightly shaded area which covers nearly a tune span of 0.5. As time goes on the en-

---

<sup>1</sup>Drawing by Regula Hofmann, Brooklyn, NY 11231, U.S.A

ergy increases and the space-charge tune spread gets smaller covering at  $t = 100$  ms the tune area shown by the darker shading. The point of highest tune corresponds to the particles which are least affected by space-charge. This point should not move much during acceleration. However, the external focusing is adjusted such that the reduced tune spread lies in about the center of the region of the tune diagram which has not to many resonances. Finally the small dark area shows the situation at  $t = 600$  ms during the acceleration when the beam has reached the top energy of 800 MeV. The reduction of the tune spread is weaker than expected from the  $1/\gamma^3$  dependence since the bunch dimensions decrease during acceleration.

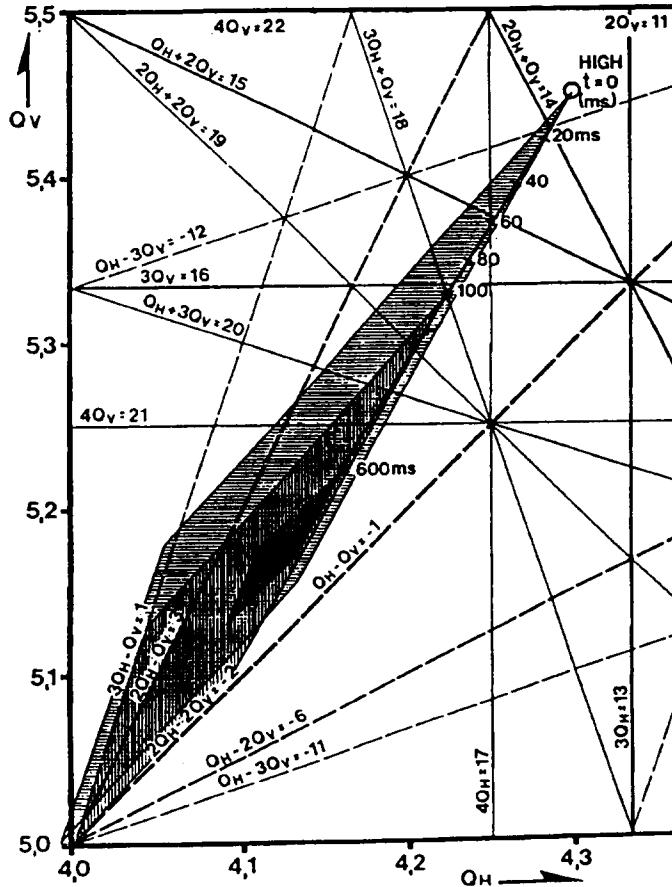


Figure 8: Direct space-charge tune shift in the PS-booster [6]

The second example shows the tune shift due to the indirect space-charge effect in the ISR [7]. The beam is unbunched and has an energy of about 26 GeV where the direct space-charge is already small due to the  $1/\gamma^2$  dependence. This beam has a large energy spread which manifests itself by the large horizontal space of nearly  $\pm 40$  mm it occupies as indicated in Fig. 9. The calculated horizontal and vertical tune shifts are shown for the different parts of the beam. They have opposite signs as expected for the indirect space-charge effect. The space occupied by this beam in the tune diagram has the form of a thin line. It can be reduced in length and straightened by some multipole magnets.

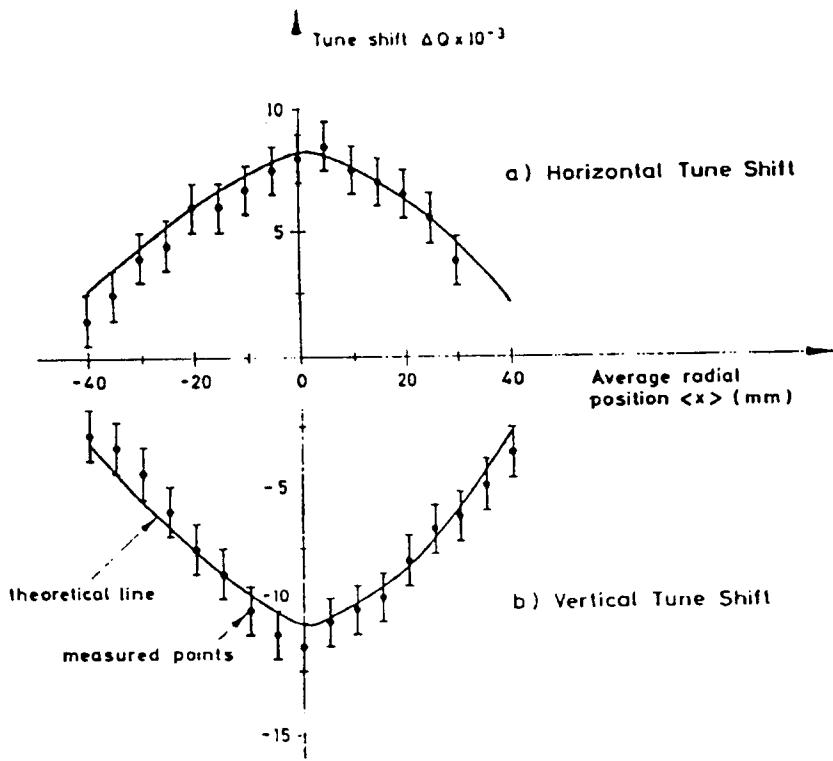


Figure 9: Indirect space-charge tune shift in the ISR [7]

## References

- [1] J. Laslett, Proc. of the 1963 Summer Study on Storage Rings, BNL-REPORT 7534, p. 324.
- [2] B. Zotter; CAS 1984 Proceedings, CERN 85-19, p. 253 (1985).
- [3] P.J. Bryant; CAS 1986 Proceedings, CERN 87-10, p. 62 (1987).
- [4] I. Hofmann, CAS 1985 Proceedings, CERN 87-03. p.327 (1987).
- [5] L. Teng, ANL-Report, ANLAD-59, (1960).
- [6] E. Brouzet, K. Schindl, CERN Booster-PS/DL/Note 80-2 (1980).
- [7] P.J. Bryant, D.M. Lewis, B. Nielson, B. Zotter, CERN ISR-MA/75-54 (1975).

# COASTING BEAM LONGITUDINAL COHERENT INSTABILITIES

*J.L. Laclare*

ESRF, Grenoble, France

## Abstract

Nowadays, the performance of most machine are limited by coherent instabilities. It is one of the most important collective effects which prevents the current from being increased above a certain threshold without the beam quality being spoiled. An intense cool beam (high intensity contained in a small phase space volume) is always unstable. A small density perturbation in the particle distribution, which can be due to either previous beam manipulations or even statistical noise (related to the point-like aspect of the particles), can grow exponentially by driving the entire beam into an unstable process. In phase space, the beam blows up, and becomes hot. With regards to the scope of the lecture, the physical mechanisms which will be considered throughout can be applied to any type of machine:

- linear accelerators,
- circular accelerators,
- storage rings

and beam:

- bunched beams
- continuous beams.

## 1. INTRODUCTION

In the first part of the lecture, we will limit ourselves to storage rings with a coasting beam which implies:

- Constant magnetic field
- no radio frequency applied.

These conditions are met in **proton or heavy-ion cooler rings**. They are also met in pulsed machines during injection or extraction of a debunched beam.

This regime does not exist on electron machines in which the RF is always on. With bunched beams, because of synchrotron motion, coherent instabilities manifest differently. However, the theoretical approach is essentially similar and easier to follow once the coasting beam case has been fully understood. The bunched beam case will be dealt with in the advanced course lecture.

The source of the instability can only be an electromagnetic field. With a very weak (very low intensity) beam, individual particles behave essentially like single particles. The external guiding field imposes the trajectories and has been designed so that these trajectories are stable.

With an intense beam, the large number of moving charges is responsible for the generation of an extra:

- "space-charge" field, or
- "self" electromagnetic field as shown in Fig. 1.

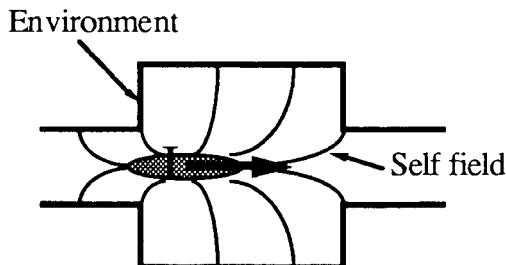


Fig. 1

If the intensity is large enough, this self field becomes sizeable in the sense that it can no longer be neglected when compared to the external guiding field. It can strongly influence the collective particle motion.

We will also limit ourselves to situations in which the external guiding field magnitude largely dominates. In other words, particle behaviour is still almost single particle behaviour. The self field acts as a perturbation.

As far as units are concerned, all formulae will be written in the International System units. Our formulae will make extensive use of the following standard quantities:

$$\begin{aligned}
 \epsilon_0 &= 8.854 \cdot 10^{-12} \text{ A s} \\
 \mu_0 &= 4 \pi \cdot 10^{-7} \text{ V s A}^{-1} \text{ m}^{-1} \\
 c &= 2.998 \cdot 10^8 \text{ m s}^{-1} \\
 \epsilon_0 \mu_0 c^2 &= 1 \\
 e &= 1.602 \cdot 10^{-19} \text{ A s}
 \end{aligned} \tag{1}$$

The self field follows Maxwell's equations.

We will do our best to avoid detailed calculations of electromagnetic fields and therefore to make explicit reference to their form. However, since these equations govern the fields and therefore the source of the instabilities, it is advisable to keep them in mind. Let  $\vec{E}$  and  $\vec{B}$  be the electric field and magnetic induction respectively. Then,  $\vec{E}$  and  $\vec{B}$  can be drawn from the potentials

$$V \text{ and } \vec{A} \text{ solutions of : } \Delta \vec{V} - \frac{1}{c^2} \frac{\partial^2 \vec{V}}{\partial t^2} = -\frac{\rho}{\epsilon_0}$$

where  $\rho$  is the particle charge density

$$\Delta \vec{A} - \frac{1}{c^2} \frac{\partial^2 \vec{A}}{\partial t^2} = -\mu_0 \vec{j} \tag{2}$$

where  $\vec{j}$  is the particle current density by means of

$$\vec{E} = -\nabla V - \frac{\partial \vec{A}}{\partial t} \text{ and } \vec{B} = \nabla \times \vec{A}$$

Obviously, boundary conditions which depend on the geometry and electromagnetic

properties of the environment (vacuum chamber, surrounding magnets, etc.) strongly influence the solution and therefore the perturbed motion of particles.

## 2. SINGLE-PARTICLE MOTION

We will first ignore the self field and consider the unperturbed single particle motion. For a coasting beam, in the first order, the longitudinal motion is very simple.

$$\frac{d\vec{p}}{dt} = e [ \vec{E} + \vec{v} \times \vec{B} ] \quad (3)$$

For the longitudinal component

$$\frac{dp_{//}}{dt} = e [ E_{//} + (\vec{v} \times \vec{B})_{//} ] \quad (4)$$

there is no longitudinal field, no RF

$$E_{//} \text{ is nul and } (\vec{v} \times \vec{B})_{//} \text{ is a second order term} \quad (5)$$

Therefore,

$$\frac{dp_{//}}{dt} = 0 \quad p_{//} \text{ is a constant of the motion} \quad (6)$$

The revolution period T around the orbit length L is written:

$$T = \frac{2\pi}{\omega} = \frac{L}{\beta c} \quad (7)$$

It depends on the particle momentum and can be expanded in terms of momentum deviation.

Let us choose a reference (machine axis for instance)

$$p_{//0} = m_0 \gamma_0 \beta_0 c \quad (8)$$

Let us then expand:

$$\beta = \beta_0 \left( 1 + \frac{1}{\gamma_0^2} \frac{p_{//} - p_{//0}}{p_{//0}} \right) \text{ and } L = L_0 \left( 1 + \alpha \frac{p_{//} - p_{//0}}{p_{//0}} + .. \right) \quad (9)$$

$$\begin{aligned} L_0 &= 2\pi R \text{ is the perimeter of the machine} \\ \alpha &= \frac{p_{//}}{L} \frac{\partial L}{\partial p_{//}} \text{ momentum compaction} \end{aligned} \quad (10)$$

in smooth machines  $\alpha \approx \frac{1}{Q_x^2}$  where  $Q_x^2$  is the horizontal betatron tune

Second- and higher-order contributions to orbit lengthening (transverse peak amplitude

$\hat{x}$  and  $\hat{z}$  dependence for instance) are neglected.

It is usual to write:

$$\frac{dT}{T} = \eta \frac{dp_{//}}{p_{//}} = - \frac{d\omega}{\omega} \quad \text{with } \eta = \alpha - \frac{1}{\gamma_0^2} = \frac{1}{\gamma_t^2} - \frac{1}{\gamma_0^2} \quad (11)$$

$$\gamma_t = \frac{1}{\alpha^{1/2}} \text{ defines the transition energy}$$

$$E_t = m_0 \gamma_t c^2 \quad (12)$$

Below transition:  $E_0 < E_t$  and  $\eta < 0$  particles with a positive momentum deviation circulate faster than the reference.

Above transition :  $E_0 > E_t$  and  $\eta > 0$  particles with a positive momentum deviation circulate slower than the reference.

When dealing with longitudinal ( $//$ ) or transverse ( $\perp$ ) instabilities the sign of  $\eta$  is essential. It indicates whether the slow particles at the tail, moving in the wake field, have a higher or lower energy than the fast particles at the front which create this wake field.

In order to describe trajectories two coordinates are necessary. Let  $p_{//0}$  and  $\omega_0$  in rad s<sup>-1</sup> be the momentum and angular revolution frequency respectively for the reference. We define two coordinates attached to the reference particle frame.

$$\tau \quad \text{and} \quad \dot{\tau} = \frac{d\tau}{dt}$$

$\tau$  is expressed in s (seconds) and represents the time delay between the passing of the reference particle and the test particle at the same point around the circumference.

$\dot{\tau}$  is the time derivation of  $\tau$ . The couple  $(\tau, \dot{\tau})$  defines the coordinates of the test particle in the longitudinal bidimensional phase space.

With the definition of  $\eta$  as given above:

$$\dot{\tau} = \frac{dT}{T} = \eta \frac{dp_{//}}{p_{//}}$$

is a constant of the motion. Accordingly, the time delay:  $\tau = \tau_0 + \dot{\tau} t$  is a linear function of time.

The differential equation of motion is:

$$\ddot{\tau} = \frac{\eta}{p_{\parallel 0}} \frac{dp_{\parallel}}{dt} = \frac{\eta e}{p_{\parallel 0}} [\vec{E} + \vec{v} \times \vec{B}]_{\parallel}(t, \theta) \quad (13)$$

Until now the right hand side has been null. Later on it will take into account the self field as a perturbation.

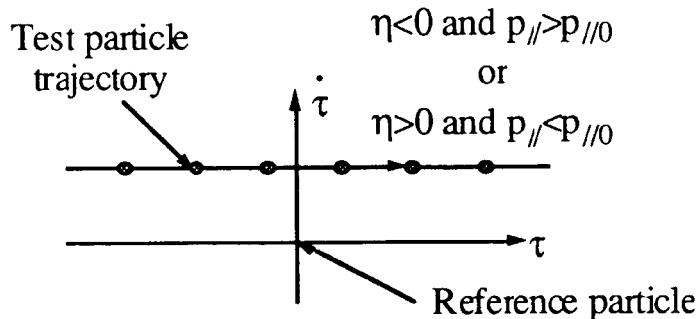


Fig. 2

### 3. LONGITUDINAL SIGNAL OF A SINGLE PARTICLE

With a view to expressing the self field, one has to solve Maxwell's equations. On the right hand side of Maxwell's equation, the expression of the beam current density  $\vec{j}(t, \theta)$  and therefore of the current which will be noted  $S/(t, \theta)$  (in Ampere) is required.

Machine physicists are used to visualizing the beam current by looking with an oscilloscope at the signal drawn from longitudinal PU electrodes. We will assume a perfect longitudinal PU electrode with infinite bandwidth located at position  $\theta$  around the ring. We will also assume that a single test particle rotates in the machine. At time  $t = 0$  the fictive reference is at  $\theta = 0$ . The PU is located at position  $\theta$  as shown in Fig. 3.

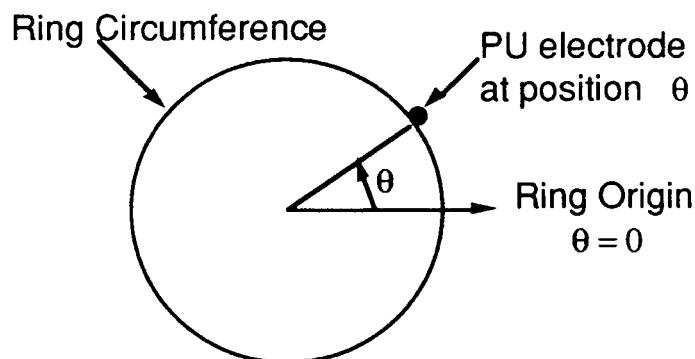


Fig. 3

The reference loops its turn with a period  $T_0$  and passes at the PU location at times:

$$t_0^p = \frac{1}{\omega_0} (\theta + 2 p \pi) \quad (14)$$

The test particle is delayed by  $\tau_0$  at time  $t = 0$ . It arrives at the PU at times:

$$t^p = t_0^p + \tau = \tau_0 + t_0^p (1 + \dot{\tau}) \quad (15)$$

The elementary signal is a series of periodical impulses delivered at each passage :

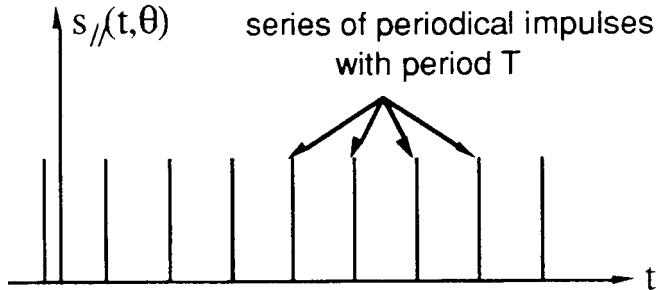


Fig. 4

The period  $T$  depends on the test particle momentum:

$$T = T_0 (1 + \dot{\tau}) = T_0 (1 + \eta) \frac{p_{\parallel} - p_{\parallel 0}}{p_{\parallel 0}} \quad (16)$$

The mathematical expression of the signal is:

$$s_{\parallel} (t, \theta) = e^{\sum_{p=-\infty}^{p=+\infty} \delta[t - \tau - \frac{1}{\omega_0} (\theta + 2 p \pi)]} \quad (17)$$

where

$$\delta[\cdot] \text{ is the Dirac function} \quad (18)$$

The series of periodical Dirac functions can be transformed into a series of exponential functions:

$$s_{\parallel} (t, \theta) = \frac{e \omega_0}{2\pi} \sum_{p=-\infty}^{p=+\infty} \exp[jp(\omega_0(t - \tau) - \theta)] \quad (19)$$

where  $j^2 = -1$

and

$$\tau = \tau_0 + \dot{\tau}t \quad (20)$$

Under this form, the signal can easily be Fourier analysed in the frequency domain.

$$\begin{aligned}
 s_{//}(\Omega, \theta) &= \frac{1}{2\pi} \int_{t=-\infty}^{t=+\infty} s_{//}(t, \theta) \exp(-j\Omega t) dt \\
 &= \frac{e\omega_0}{2\pi} \sum_{p=-\infty}^{p=+\infty} \delta\{\Omega - p\omega_0(1-\dot{\tau})\} \exp[-jp[\omega_0 \tau_0 + \theta]]
 \end{aligned} \tag{21}$$

The latter expression shows that when plugging the signal on a spectrum analyser, with  $\Omega$  along the horizontal axis, we get a series of infinitely sharp lines at all harmonics  $p\omega$  of the test particle revolution frequency:  $\omega = \omega_0(1-\dot{\tau})$  with  $p$  running from  $-\infty$  to  $+\infty$

$$\Omega_p = p\omega = p\omega_0(1-\dot{\tau}) \tag{22}$$

The spectral power amplitude is the same for all harmonics.

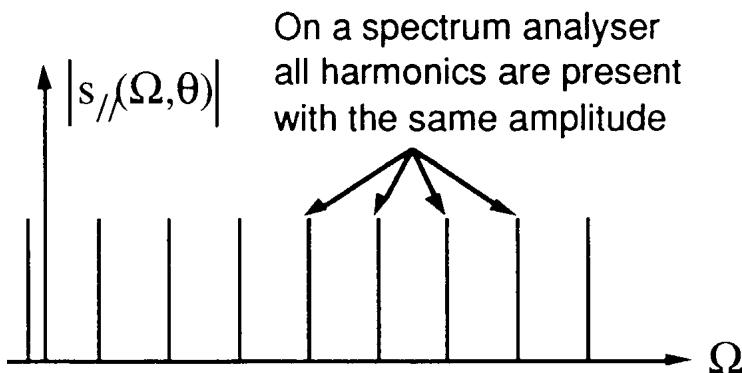


Fig. 5

If one considers a population of test particles with different momentum  $p_{//}$  and therefore  $\dot{\tau}$ , then the width of the frequency band around harmonic  $p$  is proportional to  $p$ .

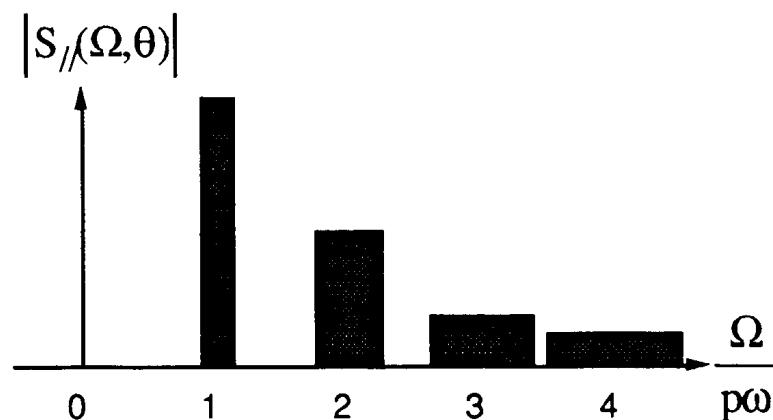


Fig. 6

In the above figure, the noise spectrum of a coasting beam with a uniform population of particles between two momenta

$$\pm \frac{\delta p_{//}}{p_{//0}} = \pm \frac{\dot{\tau}}{|\eta|} \quad (23)$$

symmetric with respect to the reference particle momentum is sketched.

One can anticipate an overlap of the frequency bands for  $|p|$  values larger than or equal to  $p_{0v}$  with:

$$\frac{1}{2p_{0v} + 1} = |\eta| \frac{\delta p_{//}}{p_{//0}} \quad (24)$$

These bands are called incoherent frequency bands.

The adjective incoherent is used to specify the fact that particles behave like independent particles, there being no coupling between them.

In accelerator physics, the electromagnetic noise picked up from the beam by means of a spectrum analyser is often called a Schottky scan. These longitudinal Schottky scans of a coasting beam make it possible to measure the momentum distribution within the beam.

#### 4. DISTRIBUTION FUNCTION

In the previous section, we obtained the expression of the current or signal of a single particle. Maxwell's equations require the total beam current. This can be obtained by adding the individual signals of all the single particles constituting the beam.

To obtain a general mathematical expression, we introduce a distribution function:  $\Psi(\tau, \dot{\tau}, t)$  which represents the particle density in the bidimensional phase space and we write the signal of the entire beam:

$$S_{//}(t, \theta) = N \int_{\tau=0}^{\tau=T_0} \int_{\dot{\tau}=-\infty}^{+\infty} \Psi(\tau, \dot{\tau}, t) s_{//}(t, \theta) d\tau d\dot{\tau} \quad (25)$$

$N$  is the total number of particles and therefore the following normalization is assumed:

$$1 = \int_{\tau=0}^{\tau=T_0} \int_{\dot{\tau}=-\infty}^{+\infty} \Psi(\tau, \dot{\tau}, t) d\tau d\dot{\tau} \quad (26)$$

Let us now introduce some basic notions and definitions concerning distributions.

A distribution is called a stationary distribution when the density does not change with time around any point of the phase space. There are as many incoming particles as departing particles around this point. We have seen that phase space trajectories are horizontal lines. Therefore, any function of  $\dot{\tau}$  only would satisfy our requirement.

On the contrary, a stationary  $\Psi$  cannot depend on  $\tau$ . As a matter of fact, a  $\tau$  dependence would automatically move to other points. Accordingly, a stationary coasting beam can be represented by:

$$\Psi(\tau, \dot{\tau}, t) = g_0(\dot{\tau}) \quad (27)$$

Such a function can represent the distribution of an unperturbed beam. It must be pointed out that the resulting signal is constant in the time domain.

$$S_{//}(t, \theta) = I \quad (28)$$

In the frequency domain there is a single line at zero frequency:

$$S_{//}(\Omega, \theta) = I \delta(\Omega) \quad (29)$$

The whole rich frequency spectrum of the individual particles composing the beam has disappeared. Such a dc current can only generate a dc electromagnetic field with essentially a transverse electric field and an azimuthal magnetic field.

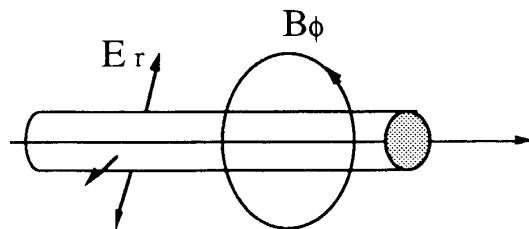


Fig. 7

The right hand side of the differential equation is null.

$$\ddot{\tau} = \frac{\eta}{p_{//0}} \frac{dp_{//}}{dt} = \frac{\eta e}{p_{//0}} [\vec{E} + \vec{v} \times \vec{B}]_{//}(t, \theta) = 0 \quad (30)$$

There is no force to drive an instability.

The conclusion is that a perfect coasting beam is always stable. In fact, a perfect coasting beam does not exist. A beam is a collection of point-like moving charges. On an average it can be represented by a stationary distribution. However, on top of it, some statistical noise is always present (Schottky scans).

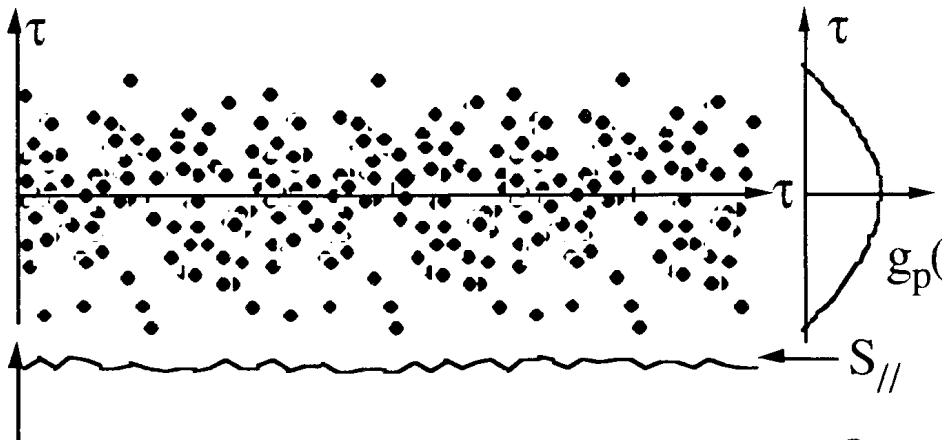


Fig. 8

When we used an integral over a distribution function (instead of a sum over individual particles) to draw the signal of the entire beam, we removed this noise. In addition, there are always distribution defects remaining as a memory of previous beam manipulations (injection kickers, remaining RF structure from a linac or a booster injector, etc...). Therefore, there are physical reasons for representing the beam by the sum of a stationary distribution plus a perturbation. Any density perturbation is automatically periodic in  $\tau$  with a period  $T_0$ . We will consider a single harmonic: harmonic  $p$  of the circumference. Accordingly, we assume a perturbation which consists of a prebunching of the beam with  $p$  wavelengths around the ring.

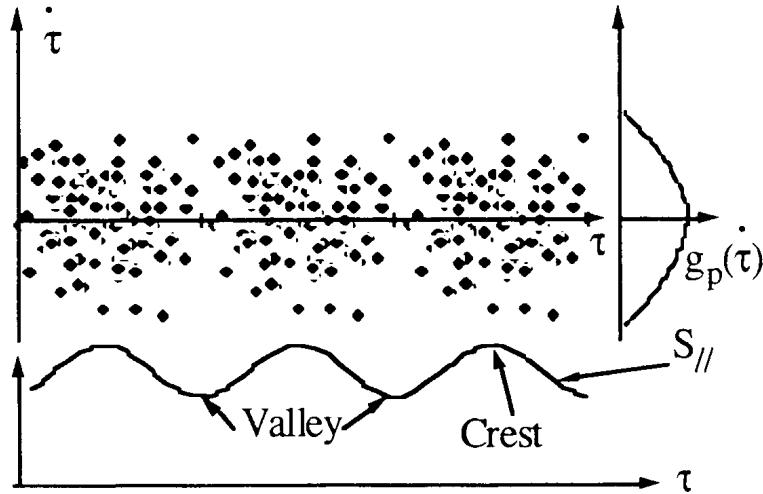


Fig. 9

This can be obtained with a term like:

$$g_p(\tau) \exp j p \omega_0 \tau \quad (31)$$

In addition, there is no reason to imagine that the perturbation will move exactly with the average particle frequency  $\omega_0$ . We will assume a slight frequency shift:  $\omega_{||}/pc$

This can be obtained by multiplying the previous term by:

$$\exp j \omega_{||}/pc t \quad (32)$$

Finally, we assume:

$$\Psi(\tau, \dot{\tau}, t) = g_0(\dot{\tau}) + g_p(\dot{\tau}) \exp j(p\omega_0\tau + \omega_{||}/pc t) \quad (33)$$

For reasons of normalization:

$$\int_{\dot{\tau}} g_0(\dot{\tau}) d\dot{\tau} = \frac{1}{T_0} \quad (34)$$

$$\int_{\tau, \dot{\tau}} g_p(\dot{\tau}) \exp j(p\omega_0\tau + \omega_{||}/pc t) d\tau d\dot{\tau} = 0 \quad (35)$$

The perturbation rearranges the particles but does not change the total number  $N$ .

Back to  $\omega_{\parallel}/pc$ , this frequency shift is a complex number.

$$\text{If } \text{Im}(\omega_{\parallel}/pc) < 0 \quad (36)$$

The perturbation increases exponentially with time. The whole beam gets bunched on harmonic p. The growth rate for self bunching is given by:

$$\frac{1}{\tau_p} = -\text{Im}(\omega_{\parallel}/pc) \quad (37)$$

On the contrary,

$$\text{If } \text{Im}(\omega_{\parallel}/pc) > 0 \quad (38)$$

then, the perturbation is damped and disappears.

The signal induced by the above distribution is given by:

$$\begin{aligned} S_{\parallel}(t, \theta) &= I + S_{\parallel p}(t, \theta) \\ &\text{with} \\ S_{\parallel p}(t, \theta) &= IT_0 \exp j[(p\omega_0 + \omega_{\parallel}/pc)t - p\theta] \int_{\tau} g_p(\tau) d\tau \end{aligned} \quad (39)$$

and Fourier analysed:

$$\begin{aligned} S_{\parallel}(\Omega, \theta) &= I \delta(\Omega) + S_{\parallel p}(\Omega, \theta) \\ &\text{with} \\ S_{\parallel p}(\Omega, \theta) &= IT_0 \exp(-jp\theta) \delta[\Omega - (p\omega_0 + \omega_{\parallel}/pc)] \int_{\tau} g_p(\tau) d\tau \end{aligned} \quad (40)$$

We find the dc component induced by the stationary distribution at zero frequency and in addition the frequency line induced by the perturbation at:

$$\Omega_C = p\omega_0 + \omega_{\parallel}/pc \quad (41)$$

This frequency line is not an incoherent frequency line but on the contrary a coherent one in the sense that we had to arrange the particles by means of a distribution to obtain this line. We have assumed an initially coherent motion. The signal or current of the beam is a complex quantity. It will be associated with complex impedances (usual practice in classical electricity).

With a view to finding a solution for  $\omega_{\parallel}/pc$ , the next step is to write down the electromagnetic field induced by the beam.

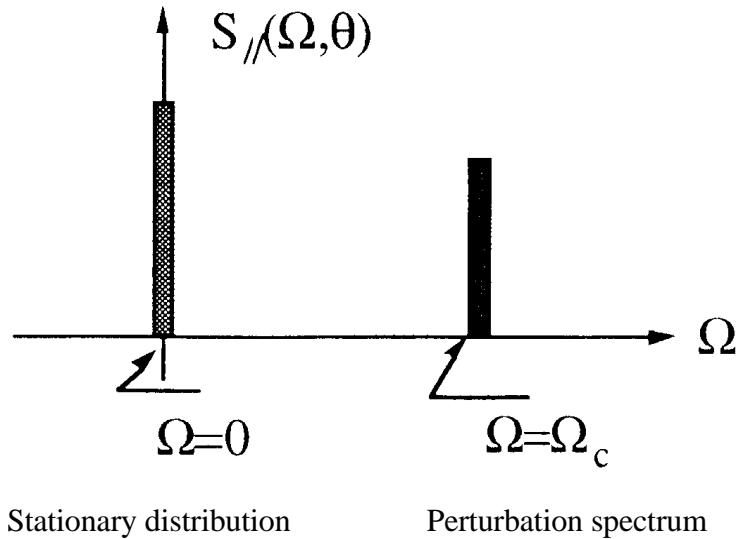


Fig. 10

## 5. ELECTROMAGNETIC FIELD INDUCED BY THE BEAM

As already stated, the solution of Maxwell's equations depends on the boundary conditions imposed by the environment. It is obvious that:

- the detailed environment seen by the beam is different for different machines and changes around the circumference of a given ring.
- one cannot expect to analytically express exact solutions for the electromagnetic field with handy formulas.

In view of this, in this chapter we will consider the most simple type of environment and use the resulting electromagnetic field expression to introduce the notion of machine coupling impedance.

Let us consider a round beam of radius  $a$  travelling in a straight line along the axis of a circular pipe of radius  $b$ . For the time being, we will also assume a perfectly conducting pipe.

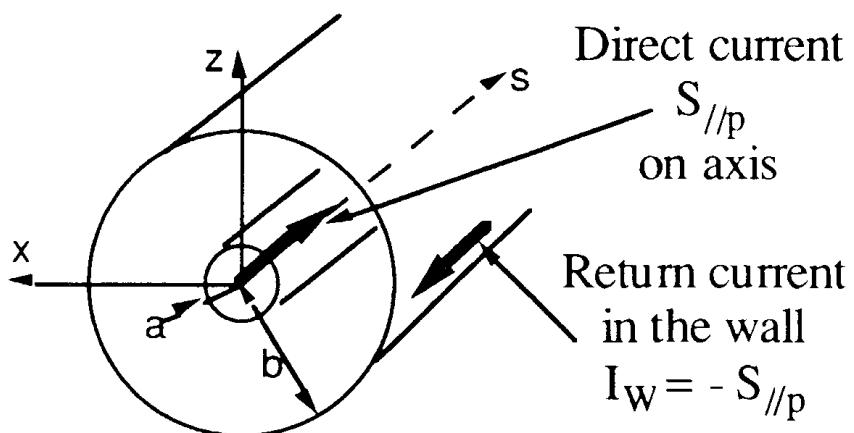


Fig. 11

With a perfectly conducting pipe, the solution of Maxwell's equations inside the pipe is independent of the environment outside the pipe. In particular the induced EM field is null outside the pipe and also in the wall thickness. It is completely stopped on the pipe inner surface by a return current  $I_w = -S_{//}/p$  uniformly distributed around the pipe, with the same amplitude as the direct signal but with the opposite sign.

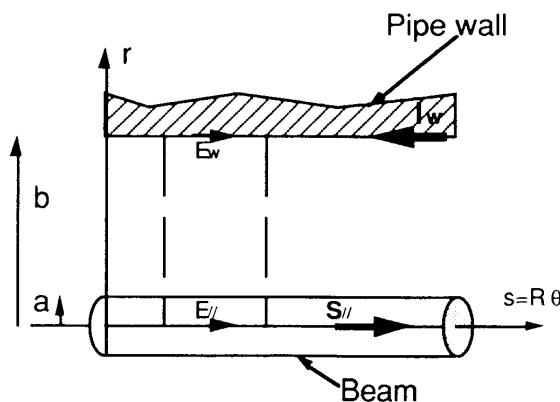


Fig. 12

Let :

$$E_{//}(t, \theta) = \widehat{E}_{//} \exp j(\Omega_C t - p\theta)$$

$$\Delta_{ab} = \frac{(K_1(z_a)I_0(z_b) + I_1(z_a)K_0(z_b))}{I_0(z_b)} \quad (42)$$

$I_n$  and  $K_n$  being  
the usual Bessel functions

Also let

$$z_r = \frac{\Omega_C r}{\beta \gamma c} \text{ for } 0 \leq r \leq a$$

Then

$$\widehat{E}_{//} = j \left( \frac{p^2}{R^2} - \frac{\Omega_C^2}{c^2} \right) \frac{\mu_0 S_{//}}{\pi \Omega_C z_a} \left[ \frac{1}{z_a} - \Delta_{ab} I_0(z_r) \right]$$

For this simple case of beam environment, the expression of the EM field is already complicated. In the longitudinal direction, the electric field is  $90^\circ$  out of phase with respect to the signal.

It is null on the crests and in the valleys of the line density when the signal is maximum, and maximum on the front and back slope when the signal is null.

Transversally, the field varies across the beam cross section:

$$I_0\left(\frac{\Omega_c r}{\beta \gamma c}\right) \text{ for } 0 \leq r \leq a \quad (43)$$

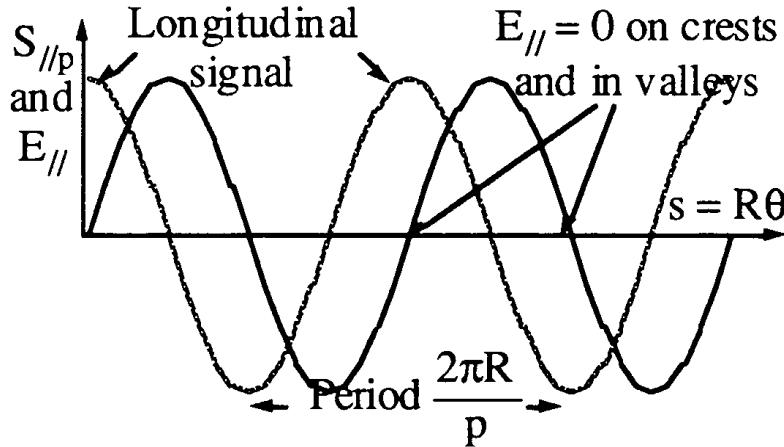


Fig. 13

This variation is small provided the argument of Bessel's function remains small at the beam edge and more generally across the pipe:  $r = b$ .

$$z_b = \frac{\Omega_c b}{\beta \gamma c} \ll 1 \quad (44)$$

where  $\Omega_c = p\omega_0 + \omega_{\parallel}/pc \approx p\omega_0 = p\frac{\beta c}{R}$

$$\text{For } z_b = 1 ; \gamma \frac{2\pi R}{p} = 2\pi b \quad (45)$$

The longitudinal wavelength with a relativistic correction  $\gamma$  in the direction of motion is equal to the pipe perimeter, or the frequency of the perturbation is equal to the pipe cut-off frequency.

$$\Omega_c = \gamma \beta \frac{c}{b} = \gamma \beta \omega_{\text{pipe cut-off}} \quad (46)$$

The EM field configuration is drastically changed around and above the cut-off frequency.

Let us evaluate orders of magnitude.

$$\begin{aligned} & \text{Assume } b = 0.03 \text{ m} \\ & \omega_{\text{pipe cut-off}} = \frac{c}{b} = 1.10^{10} \text{ rad s}^{-1} \\ & f_{\text{pipe cut-off}} = \frac{\omega_{\text{pipe cut-off}}}{2\pi} = 1.6 \text{ GHz} \end{aligned} \quad (47)$$

$$z b = \frac{\Omega_c}{\beta \gamma_{\text{pipe cut-off}}} = 1$$

implies

(48)

$$f_c(z_b=1) = \beta \gamma f_{\text{pipe cut-off}} = \beta \gamma 1.6 \text{ GHz}$$

The frequency around which this occurs depends very much on machine energy.

Consider two examples:

$$\begin{aligned} & 50 \text{ MeV/Amu } \beta=0.3 \gamma=1.05 \\ & f_c(z_b=1) = 0.5 \text{ GHz} \\ & \text{and } z_b \ll 1 \text{ leads to } f_c \leq 50 \text{ MHz} \end{aligned} \quad (49)$$

$$\begin{aligned} & 100 \text{ GeV/Amu } \beta=1 \gamma=101 \\ & f_c(z_b=1) = 160 \text{ GHz} \\ & \text{and } z_b \ll 1 \text{ leads to } f_c \leq 16 \text{ GHz} \end{aligned}$$

In the following, we will assume that we stand well below these limits, in which case a good approximation for the electric field can be obtained by using the development of Bessel's functions for small values of the argument.

$$\widehat{E_{//}(t,\theta)} = -\frac{1}{2\pi R} \frac{Z_0 g}{2j\beta_0 \gamma_0^2} \frac{\Omega_c}{\omega_0} \widehat{S_{//p}(t,\theta)}$$

where

$$Z_0 = \mu_0 c = 377 \Omega$$

$$g = 1 + 2 \operatorname{Log}\left(\frac{b}{a}\right)$$
(50)

Under all these assumptions, we have obtained the usual handy expression of the "longitudinal space charge electric field". The  $\beta_0 \gamma_0^2$  dependence shows that this field can severely affect very low energy particles.

## 6 . NEGATIVE-MASS INSTABILITY

In this section, a qualitative treatment of the longitudinal instability induced by space charge forces is presented. Immediate conclusions can be drawn from the fact that the space-charge electric field is  $90^\circ$  out of phase with respect to the signal. Then, the sign of  $\eta$

$$\frac{d\omega}{\omega_0} = -\eta \frac{dp}{p_{//0}} \text{ with } \eta = \frac{1}{\gamma_t^2} - \frac{1}{\gamma_0^2} \quad (51)$$

defines whether the perturbation amplitude increases (instability) or decreases (stability). Below transition energy ( $\eta < 0$ )

- acceleration on the front slope (2 in Fig. 14) means increased momentum and therefore higher revolution frequency. Particles move ahead.

- deceleration on the back slope (3) implies that particles will move backwards.

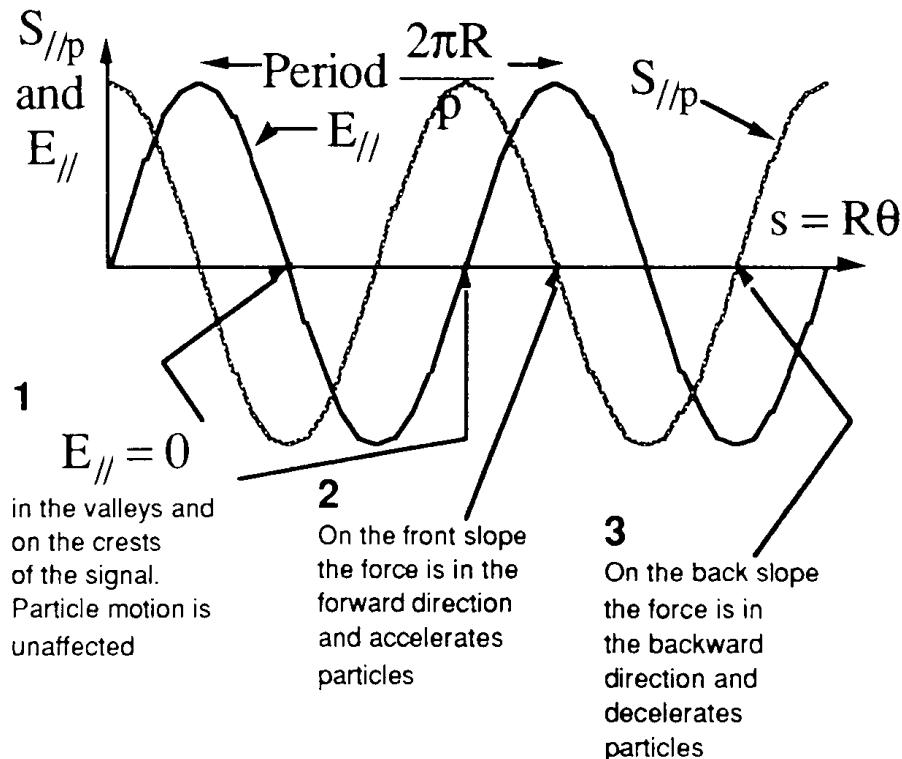


Fig. 14

Conclusion: Particles will move from the crests to fill the valleys and the initial perturbation will disappear.

It can be concluded that space charge has a stabilizing effect below transition energy. Obviously, the reverse applies above transition when  $\eta > 0$ . Above transition, space-charge fields make the coasting beam potentially unstable. This effect is commonly called negative-mass instability in view of the apparently strange fact that above transition, particles with higher energy go slower in the machine (longer trajectory even though their velocity is higher).

## 7. INTRODUCTION OF THE LONGITUDINAL COUPLING IMPEDANCE $Z_{\parallel}(\omega)$

At this point, we could have started a quantitative treatment of the detailed motion of the perturbation. However, our pipe model of beam environment is very crude and our conclusions regarding stability would not be relevant to an actual machine. On the other hand, our mathematical possibilities are very limited. The analytical solution for the simple pipe model was complicated already. Extracting complete solutions of Maxwell's equations taking due account of an actual machine environment (all the changes of the geometry and electromagnetic properties of the vacuum chamber around the circumference) is definitely impossible. In order to remain as general as possible in our discussion about stability we shall introduce a new machine parameter: the longitudinal coupling impedance  $Z_{\parallel}(\omega)$

$Z_{\parallel}(\omega)$  is in fact a parameter which can only be drawn once Maxwell's equations have

been completely solved. It gathers all the details of the electromagnetic coupling between the beam and the surroundings. By varying its form, we will be able to consider any type of beam self electromagnetic field.

To introduce this parameter, let us use the expression of the space charge field. This expression shows that the field is proportional to current.

$$\widehat{E_{//}}(t,\theta) = -\frac{1}{2\pi R} \left[ \frac{Z_0 g}{2j\beta_0 \gamma_0^2} \frac{\Omega_c}{\omega_0} \right] \widehat{S_{//p}}(t,\theta) \quad (52)$$

It is written in a form identical to Ohm's law and the quantity between brackets must be expressed in  $\Omega$ . This latter quantity will be called the longitudinal coupling impedance in general:

$$\widehat{E_{//}}(t,\theta) = -\frac{1}{2\pi R} Z_{//(w)} \widehat{S_{//}}(t,\theta) \quad (53)$$

and longitudinal space charge impedance in the present particular case with:

$$Z_{//SC}(w) = \frac{Z_0 g}{2j\beta_0 \gamma_0^2} \frac{\omega}{\omega_0} \quad (54)$$

At this stage, some generalization is necessary. Since both electric and magnetic forces act on the particles:

$$\ddot{\tau} = \frac{\eta}{p_{//0}} \frac{dp_{//}}{dt} = \frac{\eta e}{p_{//0}} [\vec{E} + \vec{v} \times \vec{B}]_{//}(t,\theta) \quad (55)$$

we include the magnetic field contribution in our definition and write:

$$[\vec{E} + \vec{v} \times \vec{B}]_{//}(t,\theta) = -\frac{1}{2\pi R} Z_{//(w)} S_{//}(t,\theta) \quad (56)$$

Up to now we have considered a pure perturbation at a single frequency. However, the most general beam signal spectrum (bunched beam for instance) is not limited to a single frequency but is spread over a wide frequency range.

Each individual frequency  $\omega$  present in the Fourier transform of the signal contributes to the force and must be combined with the corresponding impedance at  $\omega$ .

$$[\vec{E} + \vec{v} \times \vec{B}]_{//(t,\theta)} = \frac{-1}{2\pi R} \int_{\omega=-\infty}^{\omega=+\infty} Z_{//(w)} S_{//(w,\theta)} \exp j\omega t d\omega \quad (57)$$

in which the following definitions are assumed.

$$S_{//}(\omega, \theta) = \frac{1}{2\pi} \int_{t = -\infty}^{t = +\infty} S_{//}(t, \theta) \exp(-j\omega t) dt$$

and reciprocally

(58)

$$S_{//}(t, \theta) = \int_{\omega = -\infty}^{\omega = +\infty} S_{//}(\omega, \theta) \exp(j\omega t) d\omega$$

With a view to obtaining an expression of  $Z_{//}(\omega)$  in terms of physical quantities, let us apply the previous relation to a Dirac charge  $e$  rotating at frequency  $\omega_0$  in the machine.

$$S_{//}(t, \theta) = \sum_p e \delta(t - \frac{\theta + 2k\pi}{\omega_0})$$

$$S_{//}(\omega, \theta) = \frac{e\omega_0}{2\pi} \sum_p \exp(-jp\theta) \delta(\omega - p\omega_0)$$

$$-2\pi R[\vec{E} + \vec{v} \times \vec{B}]_{//}(t, \theta) = \frac{e\omega_0}{2\pi} \sum_p Z_{//}(p\omega_0) \exp(jp(\omega_0 t - \theta))$$
(59)

We now follow the field particle while remaining  $\tau$  behind it by taking  $\theta = \omega_0(t - \tau)$  and obtain

$$-\frac{2\pi R}{e} [\vec{E} + \vec{v} \times \vec{B}]_{//}(\tau) = G(\tau)$$

$$= \frac{\omega_0}{2\pi} \sum_p Z_{//}(p\omega_0) \exp(jp\omega_0\tau)$$
(60)

The voltage per unit charge  $G(\tau)$  expressed in volts per Coulomb is usually called the **Green function**. It is expanded into a series over all harmonics of the revolution frequency.

If the impedance is smooth enough or the machine long enough so that the wake field is null after one revolution, the series can be approximated by an integral.

$$G(\tau) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} Z_{//}(\omega) \exp(j\omega\tau) d\omega$$
(61)

This last expression can be inverted to obtain

$$Z_{//}(\omega) = \int_{-\infty}^{+\infty} G(\tau) \exp(-j\omega\tau) d\tau \quad (62)$$

In other words,  $Z_{//}(\omega)$  is the Fourier transform of  $2\pi G(\tau)$ .

Ideally, this relation can be used in numerical codes to calculate the function  $Z_{//}(\omega)$ . A short bunch of particles is sent on axis through a structure. Maxwell's equations are solved step by step at predetermined mesh points. This allows the Green function and afterwards its Fourier transform to be obtained numerically.

## 8 . LONGITUDINAL COUPLING IMPEDANCE $[Z_{//}(\omega)]$ OF AN ACCELERATOR RING

Designers are confronted with the problem that they have to include in their machine all sorts of bellows, flanges, cross section changes, PU electrodes, RF gaps, kickers, septum magnets, etc., which influence the impedance. During the last two decades, progress has been made towards estimating impedances better. Numerical codes and experimental tools have been developed. However, RF properties are difficult to predict or measure and we are still far from the situation where we could predict the curve  $Z_{//}(\omega)$  accurately enough before commissioning a machine. We have also learnt a lot from existing machines. Most of them suffering from instabilities despite many attempts to measure and lower their impedance.

As already stated, the impedance is a complex function of frequency. A priori, it has an imaginary part and a real part. The space charge impedance from a perfectly conducting round pipe was an example of a purely imaginary impedance. Basically, in the broad sense, a ring impedance can be inductive, or capacitive or resistive. Our equations are written in such a form that we have to consider the complete frequency axis (positive and negative  $\omega$ ). In addition, for reasons which will be understood later in the course of this lecture, the quantity of interest is not directly

$$Z_{//}(\omega) \text{ but } \frac{Z_{//}(\omega)}{\omega} \text{ or } \frac{Z_{//}(\omega)}{p} \text{ with } p = \frac{\omega}{\omega_0} \quad (63)$$

Accelerator physicists are used to plotting

$$\text{Im}\left(\frac{Z_{//}(\omega)}{\omega}\right) \text{ and } \text{Re}\left(\frac{Z_{//}(\omega)}{\omega}\right) \quad (64)$$

with  $\omega$  along the horizontal axis.

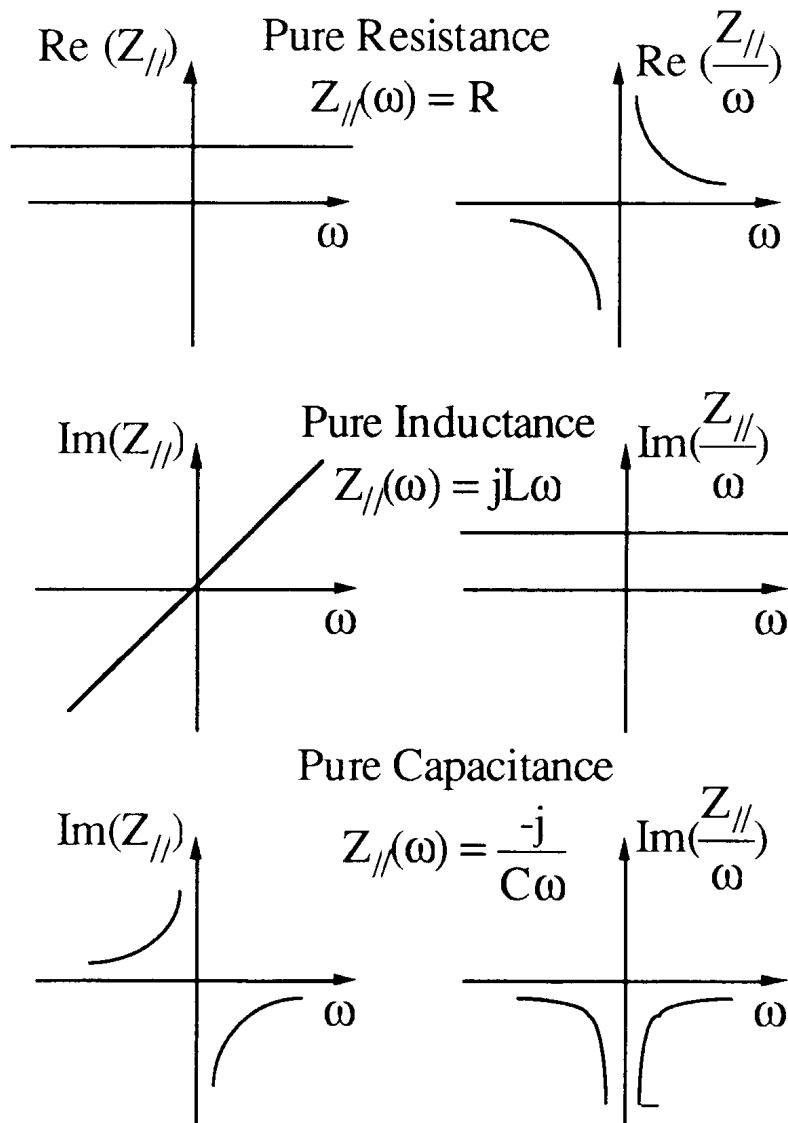


Fig. 15

In a standard machine, out of the four major components of the impedance, the first one is the space charge component already introduced.

$$\frac{Z_{\parallel/SC}(\omega)}{p} = -j \frac{Z_0 g}{2\beta_0 \gamma_0^2} \text{ where } p \text{ stands for } \frac{\omega}{\omega_0} \quad (65)$$

It corresponds to a pure negative inductance (imaginary part only) very large for low energies. For instance:

$$50 \text{ MeV/Amu } \beta_0=0.3 \gamma_0=1.05 g \approx 2.4 \frac{Z_{\parallel/SC}}{p} = -j1.4 \text{ k}\Omega \quad (66)$$

$$10 \text{ GeV/Amu } \beta_0=1 \gamma_0=11.7 g \approx 2.4 \frac{Z_{\parallel/SC}}{p} = -j3.3 \Omega$$

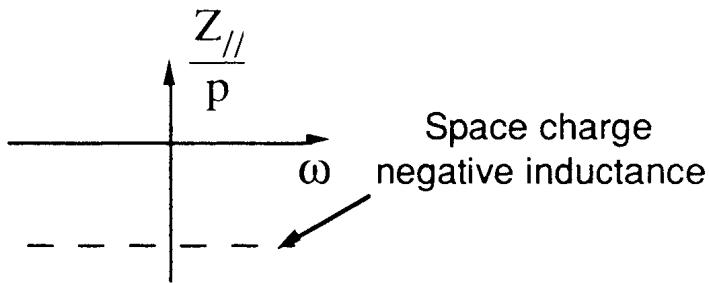


Fig. 16

The resistive wall impedance is the second major component.

Previously, the pipe was assumed to have a perfectly conducting wall (conductivity  $\sigma = \infty$ ) and no resistance was opposed to the return wall current  $I_w$ . For finite  $\sigma$ ,  $I_w$  flows in a strip of length  $2\pi R$ , width  $2\pi b$ , thickness  $\delta$ : the skin depth at frequency  $\omega$ . On top of the space-charge field, the beam sees a uniform longitudinal electric field set up by  $I_w$ .  $\delta_0$  being the skin depth at frequency  $\omega_0$ . We will write:

$$\delta^2 = \delta_0^2 \frac{\omega_0}{\omega} \text{ with } \delta_0^2 = \frac{2}{\mu\sigma\omega_0} \quad (67)$$

and consider two regimes.

At very low frequency, the skin depth  $\delta$  is larger than the wall thickness  $\delta_w$  (thin wall). The impedance seen by  $I_w$  is the pipe resistance

$$Z_{//RW} = \frac{1}{\sigma} \frac{2\pi R}{2\pi b \delta_w} \quad \text{or} \quad \frac{Z_{//RW}}{p} = \frac{R}{\sigma b \delta_w} \frac{1}{p} \quad (68)$$

The impedance seen by the beam is exactly the same.

At high frequencies, the wall is thicker than the skin depth. It can be shown that the previous formula must be amended by replacing  $\delta_w$  by  $\delta$  and multiplying by  $(1+j)$  (an imaginary term appears).

$$\frac{Z_{//RW}}{p} = (1+j) \frac{R}{\sigma b \delta} \frac{\omega_0}{\omega} = (1+j) \frac{Z_0 \beta_0 \delta_0}{2b} \frac{1}{\sqrt{p}} \quad (69)$$

The transition between the two expressions occurs when  $\delta = \delta_w$ .

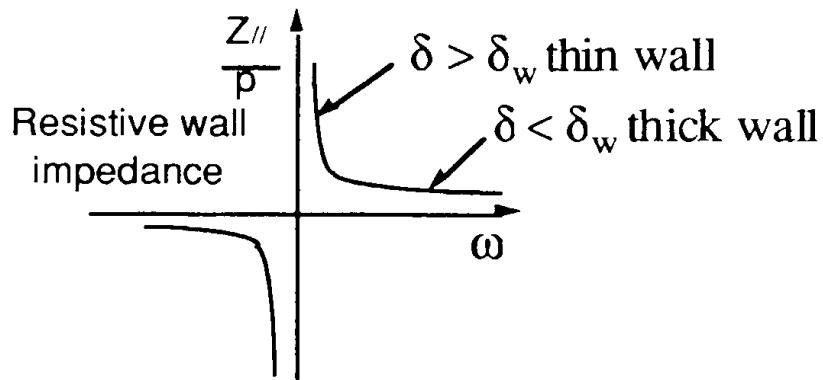


Fig. 17

The energy lost in the wall is drawn from the beam which is decelerated. As far as instabilities are concerned, the resistive wall impedance is not the source of problems in the longitudinal direction. It essentially affects the transverse motion.

The two last major components correspond to resonating objects. Let us start by some general definitions and remarks concerning resonators. The impedance of a resonator is often written:

$$Z_{//} = \frac{R_s}{1 + jQ\left(\frac{\omega}{\omega_r} - \frac{\omega_r}{\omega}\right)} \quad (70)$$

$R_s$  is the shunt impedance in  $\Omega$ ;  $Q$  the quality factor ;  $\omega_r$  the resonant frequency

When the resonator is driven at very low frequency,

$$\omega = p\omega_0 \ll \omega_r \quad \frac{Z_{//}}{p} = j \frac{R_s \omega_0}{Q \omega_r} \quad (71)$$

it behaves like a pure inductance. At resonance, it is a pure resistance.

$$\omega = p\omega_0 = \omega_r \quad \frac{Z_{//}}{p} = R_s \frac{\omega_0}{\omega_r} \quad (72)$$

At high frequencies,

$$\omega = p\omega_0 \gg \omega_r \quad \frac{Z_{//}}{p} = R_s \frac{\omega_0 \omega_r}{jQ\omega^2} \quad (73)$$

it behaves like a pure capacitance.

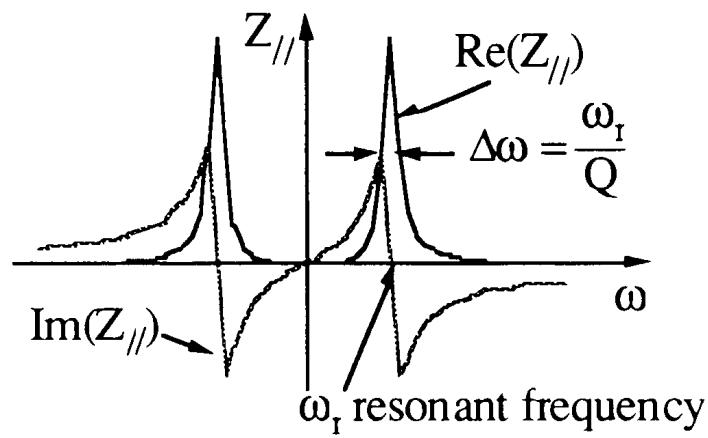


Fig. 18

As shown in Fig. 18, the real part is maximum ( $R_s$ ) at the resonant frequency. The resonator bandwidth  $\Delta\omega$  full width, half height is given by:

$$\Delta\omega = \frac{\omega_r}{Q}.$$

It can be shown that the time  $\Delta t$  during which a resonator of bandwidth  $\Delta\omega$  memorizes the energy left by a Dirac excitation is given by:

$$\Delta\omega\Delta t = \frac{\omega_r}{Q} \Delta t \approx 1 \quad (74)$$

(uncertainty principle).

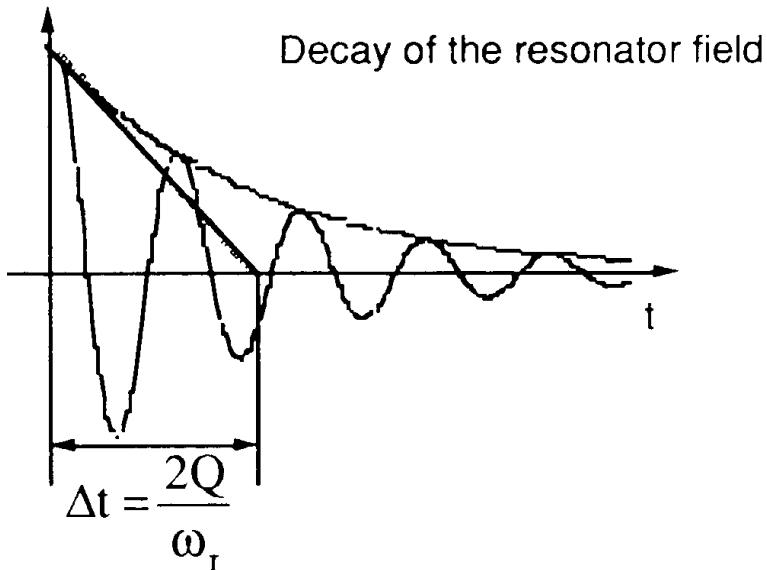


Fig. 19

The immediate consequence is that the response or wake field of a narrow-band resonator

(high Q and small  $\Delta\omega$ ) lasts for a long time. Particles well separated in time are coupled by such a resonant object. The RF cavities are the most current sources of narrow-band resonators. They are tuned to resonate at the fundamental frequency  $\hbar\omega_0$ . However, resonant higher-order longitudinal (and also transverse) modes with high Q are always present. The representation of a narrow-band resonator in the impedance diagram with  $\frac{Z_{||}}{p}$  along the vertical axis is sketched below.

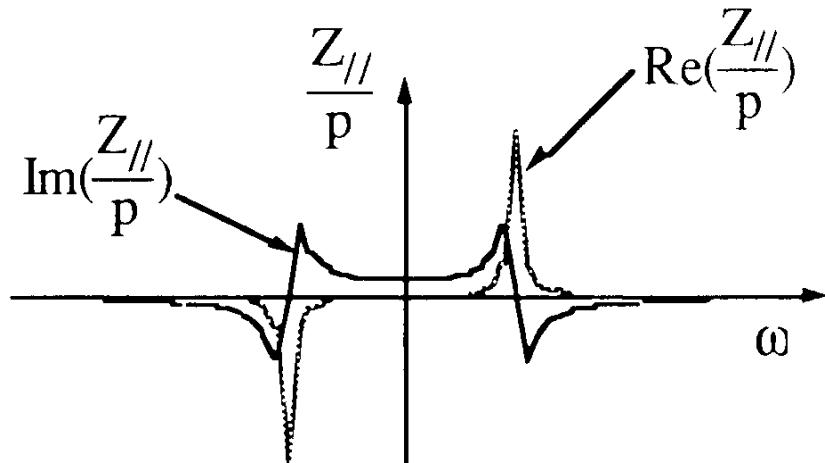


Fig. 20

Narrow band resonators constitute the third type of impedance usually met in an accelerator ring. The last component of the impedance corresponds to the numerous changes of cross sections, bellows, flanges, etc. It is obvious that these structures can trap some magnetic field and therefore behave like an inductance at low frequencies. This has been measured on existing machines.

We also learnt that when no special care is taken, the vacuum chamber is essentially resistive at frequencies around the pipe cut-off frequency. This is due to the fact that the path followed by the return current is very complicated and the resistance high when the vacuum chamber wall is not smooth or correctly shielded along the longitudinal axis. It has also been observed that the resistive part drops at frequencies higher than the cut-off frequency.

The object is to represent the above observations with the simplest impedance model. In this respect, a broad band resonator with a resonant frequency around the vacuum pipe cut-off frequency:

$$\omega_{rBB} \approx \omega_{cut-off} = \frac{c}{b} \quad (75)$$

can give an overall satisfactory result. To a certain extent, most experimental results drawn from existing rings have been correctly fitted by assuming the existence of such a component with  $Q \approx 1$  (as sketched below) and a shunt impedance  $R_s$  adjusted to obtain the good value of the low frequency inductance.

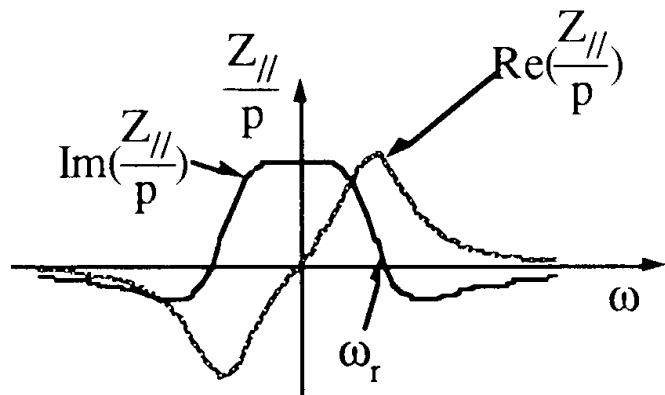


Fig. 21

Concerning orders of magnitudes, the full range

$$0.2 \Omega \leq \left| \frac{Z_{//}}{p} \right|_{\max} = R_s \frac{\omega_0}{\omega_r} \leq 50 \Omega \quad (76)$$

has been found. The lowest values are obtained with modern machines. At present, a considerable effort is being put into designing a very smooth chamber. Unavoidable changes of cross section are systematically shielded and are no longer seen by the beam.

For such low-Q objects, the impedance curve varies slowly with frequency and the resonator bandwidth  $\Delta\omega$  is large. Therefore, the wake field decays rapidly. It is a local interaction which can only couple particles close to each other along the longitudinal axis.

This handy broad band-model can only be used as a crude representation of reality. It will constitute the last component of our inventory of contributions to a ring impedance.

## 9 . VLASOV'S EQUATION AND DISPERSION RELATION

At this point, all necessary ingredients to start the most general discussion on beam stability have been gathered.

– distribution function:  $\Psi(\tau, \dot{\tau}, t) = g_0(\dot{\tau}) + g_p(\dot{\tau}) \exp j(p\omega_0\tau + \omega_{//}pc t)$  (77)

– perturbation signal:  $S_{//p}(t, \theta) = IT_0 \exp j[(p\omega_0 + \omega_{//}pc)t - p\theta] \int_{\dot{\tau}} g_p(\dot{\tau}) d\dot{\tau}$  (78)

– the general impedance of a ring  $Z_{//}$ ,

– the general expression of the EM field generated by a signal in a ring with impedance  $Z_{//}$

$$[\vec{E} + \vec{v} \times \vec{B}]_{//}(t, \theta) = \frac{-1}{2\pi R} \int_{\omega=-\infty}^{\omega=+\infty} Z_{//}(\omega) S_{//}(\omega, \theta) \exp j\omega t d\omega \quad (79)$$

$$[\vec{E} + \vec{v} \times \vec{B}]_{//}(t, \theta) = -\frac{1}{2\pi R} Z_{//}(\Omega_C) S_{//}(t, \theta) \quad (80)$$

in the present case.

$$= -\frac{I}{\beta_0 c} Z_{//}(\Omega_C) \exp j[\Omega_C t - p\theta] \int_{\tau} g_p(\dot{\tau}) d\dot{\tau} \quad (81)$$

where  $\Omega_C$  stands for  $p\omega_0 + \omega_{//}/pc$

when the expression of the perturbation signal is used.

– The EM force acts back on the particles,

$$\ddot{\tau} = \frac{\eta}{p_{//0}} \frac{dp_{//}}{dt} = \frac{\eta e}{p_{//0}} [\vec{E} + \vec{v} \times \vec{B}]_{//}(t, \theta) \quad (82)$$

in which  $\theta = \omega_0(t - \tau)$  in order to follow the particle along its trajectory.

$$\ddot{\tau} = \frac{-\eta e I Z_{//}(\Omega_C)}{m_0 \gamma_0 c^2 \beta_0^2} \exp j[p\omega_0 t + \omega_{//} c t] \int_{\tau} g_p(\dot{\tau}) d\dot{\tau} \quad (83)$$

The goal is to find out the solution

$$\Omega_C \text{ or } \omega_{//}/pc \quad (84)$$

and to discuss beam stability via the imaginary part. The evolution of the distribution is governed by the collision-free Boltzmann equation:

$$\frac{\partial \Psi}{\partial t} + \text{div}(\Psi \vec{v}) = 0 \quad (85)$$

This equation is similar to the one commonly used in electromagnetism for the charge density

$$\frac{\partial \rho}{\partial t} + \text{div} \vec{j} = 0 \text{ with } \vec{j} = \rho \vec{v} \quad (86)$$

In our particular case:

$$\vec{v} = \begin{pmatrix} \frac{d\tau}{dt} \\ \vdots \\ \frac{d\tau}{dt} \end{pmatrix} \quad (87)$$

The equation can be developed  $\frac{\partial \Psi}{\partial t} + \vec{v} \cdot \vec{\nabla} \Psi + \Psi \text{ div } \vec{v} = 0$  (88)

and then simplified, since  $\operatorname{div} \vec{v} = 0$  when canonical variables like  $\tau, \dot{\tau}$  are used.

$$\frac{\partial \Psi}{\partial t} + \vec{v} \cdot \vec{\operatorname{grad}} \Psi = 0$$

$$\frac{\partial \Psi}{\partial t} + \frac{\partial \Psi}{\partial \tau} \dot{\tau} + \frac{\partial \Psi}{\partial \dot{\tau}} \ddot{\tau} = \frac{d\Psi}{dt} = 0 \quad (89)$$

In this form, it is called **Vlasov's equation**.

We now apply Vlasov's equation to the assumed distribution :

$$\Psi(\tau, \dot{\tau}, t) = g_0(\dot{\tau}) + g_p(\dot{\tau}) \exp j(p\omega_0\tau + \omega_{||}/pc^t) \quad (90)$$

and get:

$$\begin{aligned} \frac{\partial \Psi}{\partial t} &= j\omega_{||}/pc g_p(\dot{\tau}) \exp j(p\omega_0\tau + \omega_{||}/pc^t) \\ \dot{\tau} \frac{\partial \Psi}{\partial \tau} &= j p\omega_0 \dot{\tau} g_p(\dot{\tau}) \exp j(p\omega_0\tau + \omega_{||}/pc^t) \\ \ddot{\tau} \frac{\partial \Psi}{\partial \dot{\tau}} &= \ddot{\tau} \frac{\partial g_0}{\partial \dot{\tau}} + \text{second order term} \frac{\partial g_p}{\partial \dot{\tau}} \ll \frac{\partial g_0}{\partial \dot{\tau}} \end{aligned} \quad (91)$$

Limiting ourselves to first order terms, we obtain:

$$g_p(\dot{\tau}) = j \frac{\ddot{\tau}}{p\omega_0 \dot{\tau} + \omega_{||}/pc} \frac{\partial g_0}{\partial \dot{\tau}} \quad (92)$$

in which

$$\ddot{\tau} = \frac{-\eta I Z_{||}/(\Omega_c)}{(\frac{m_0 c^2}{e}) \gamma_0 \beta_0^2} \exp j[p\omega_0 \tau + \omega_{||}/pc^t] \int_{\dot{\tau}} g_p(\dot{\tau}) d\dot{\tau} \quad (93)$$

At first glance, it seems difficult to exploit this expression since the function  $g_p$  which gives the details of the initial perturbation appears explicitly. However, one can integrate both sides of the equation over  $\dot{\tau}$  values. The equation then obtained from which the solution  $\Omega_c$  will be drawn is called the **Dispersion Relation**. It is independent of  $g_p$ . In other words, the growth rate of the instability is independent of the exact form of  $g_p$ .

$$1 = \frac{\eta(\frac{q}{A})I}{(\frac{m_0 c^2}{e}) \gamma_0 \beta_0^2} j \frac{Z_{||}/(\Omega_c)}{p\omega_0} \int_{\dot{\tau}} \frac{\frac{\partial g_0}{\partial \dot{\tau}}}{\frac{\omega_{||}/pc}{p\omega_0} + \dot{\tau}} d\dot{\tau} \quad (94)$$

The Dispersion Relation has been written for all types of particles :

$$\begin{aligned}
 \frac{q}{A} &= 1 \text{ for protons or electrons} \\
 q &= \text{number of charges} \\
 A &= \text{number of masses} \\
 &\quad \text{for heavy ions} \\
 &\quad 0.511 \text{ MV for electrons} \\
 \frac{m_0 c^2}{e} &= 0.938 \text{ GV for protons} \\
 &\quad 0.932 \text{ GV for heavy ions}
 \end{aligned} \tag{95}$$

A priori, the frequency shift  $\omega_{\parallel}/pc$  appears twice,

$$1 = \frac{-\eta(\frac{q}{A})I}{(\frac{m_0 c^2}{e})\gamma_0 \beta_0^2} j \frac{Z_{\parallel}(\Omega_C)}{p\omega_0} \int \frac{\frac{\partial g_0}{\partial \tau}}{\frac{\omega_{\parallel}/pc}{p\omega_0} + \tau} d\tau \tag{96}$$

under the integral sign and as an argument of

$$Z_{\parallel}(\Omega_C) = Z_{\parallel}(p\omega_0 + \omega_{\parallel}/pc) \tag{97}$$

However, it is assumed that

$$\omega_{\parallel}/pc \ll p\omega_0 \tag{98}$$

and therefore, the value of the impedance can be taken at  $p\omega_0$ . Accordingly, we will write:

$$Z_{\parallel}(\Omega_C) = Z_{\parallel}(p) \tag{99}$$

We understand now why the quantity of interest is not

$$Z_{\parallel}(p) \text{ but } \frac{Z_{\parallel}(p)}{p} \tag{100}$$

To avoid carrying heavy expressions we will use the quantity:

$$\Lambda_{//} = \frac{\eta(\frac{q}{A})I}{2\pi \left(\frac{m_0 c^2}{e}\right) \gamma_0 \beta_0^2} \quad (101)$$

and write the Dispersion Relation

$$1 = -\frac{2\pi \Lambda_{//}}{\omega_0} j \frac{Z_{//(p)}}{p} \int \frac{\frac{\partial g_0}{\partial \dot{\tau}}}{\frac{\omega_{//} pc}{p\omega_0} + \dot{\tau}} d\dot{\tau} \quad (102)$$

## 10. MONOCHROMATIC BEAM

We apply the Dispersion Relation

$$1 = -\frac{2\pi \Lambda_{//}}{\omega_0} j \frac{Z_{//(p)}}{p} \int \frac{\frac{\partial g_0}{\partial \dot{\tau}}}{\frac{\omega_{//} pc}{p\omega_0} + \dot{\tau}} d\dot{\tau} \quad (103)$$

to an infinitely cool beam with momentum  $p_{//0}$ :

$$g_0(\dot{\tau}) = \delta(\dot{\tau})$$

$$\int \frac{\frac{\partial g_0}{\partial \dot{\tau}}}{\frac{\omega_{//} pc}{p\omega_0} + \dot{\tau}} d\dot{\tau} = \frac{\omega_0}{2\pi} \left( \frac{\omega_{//} pc}{p\omega_0} \right)^{-2} \quad (104)$$

Finally one obtains:

$$\left( \frac{\omega_{//} pc}{p\omega_0} \right)^2 = -\frac{\Lambda_{//}}{\eta} j \eta \frac{Z_{//(p)}}{p} \text{ with } \frac{\Lambda_{//}}{\eta} \text{ positive} \quad (105)$$

- 1) Let us first consider a pure resistive component. The equation has two roots and whatever the sign of  $p$ , one of the two roots has  $\text{Im}(\omega_{//} pc) < 0$ . The beam is unstable.

2) We assume now a pure space-charge interaction.

$$j \frac{Z_{//} SC(p)}{p} = \frac{Z_0 g}{2\beta_0 \gamma_0^2} \text{ is real and positive} \quad (106)$$

- below the transition energy ( $\eta < 0$ ) the beam is stable.
- above the transition energy ( $\eta > 0$ ) the beam is unstable (negative-mass instability).

3) We now consider the effect of a pure inductance, for instance the broad-band impedance at low frequencies. The results obtained with space charge must be inverted. In particular the beam is stable above transition.

The mathematical tools used to solve our problem (Vlasov's equation, Dispersion Relation, complex plane for frequencies and impedances, etc.) are very powerful. However, the physical meaning is rather lost in this series of mathematical treatments. We have begun with a prebunching of the beam with  $p$  wavelengths around the circumference. This prebunching coupled with the impedance creates a longitudinal force at a frequency close to  $p\omega_0$  which reacts back on particles like an RF cavity would do. As a matter of fact the expression of  $\omega_{//pc}$  is in all respects identical to the standard expression of the synchrotron frequency  $\omega_s$  on a flat top of a magnetic field.

$$\omega_s^2 = \frac{-\eta(\frac{q}{A}) \widehat{V_{RF} \cos \phi_s} h\omega_0^2}{2\pi \frac{m_0 c^2}{e} \gamma_0 \beta_0^2} \quad \frac{\omega_{//pc}^2}{\omega_s^2} = \frac{[jZ_{//}(p)I]}{\widehat{V_{RF} \cos \phi_s}} \quad (107)$$

$$\widehat{V_{RF} \cos \phi_s} \text{ is replaced by } [jZ_{//}(p)I] \quad (108)$$

The results (curves corresponding to a given growth rate  $\text{Im}(\omega_{//pc})$ ), are generally represented in the impedance complex plane as shown below.

With the exception of the vertical axis (pure space charge below transition and broad band inductance above transition), all the working points in the impedance diagram correspond to an unstable motion. For a given impedance, the Dispersion Relation tells us at which rate a perturbation will grow. This is certainly the essential piece of information. However, the details of the evolution of the beam deterioration are inaccessible.

Let us qualitatively describe the instability in phase space. With a working point along the vertical axis in the unstable region (point A, Fig. 22), the resistive part of the impedance is null. The frequency shift  $\omega_{//pc}$  has no real part which means that the RF field generated by the perturbation is exactly tuned to  $p\omega_0$ . The peaks of longitudinal density get trapped in the center of the buckets as shown in Fig. 23. The growth time is the synchrotron period in these bucket

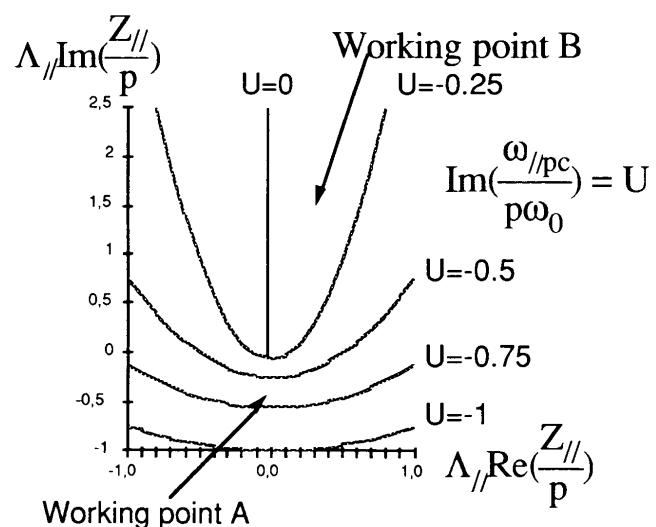


Fig. 22

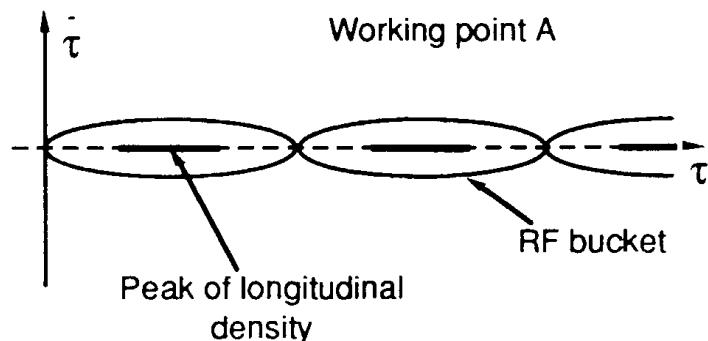


Fig. 23

When the working point is associated with a resistance (point B, Fig. 22), the frequency shift has both a real and an imaginary part. The RF buckets are generated at

$$\Omega_C = p\omega_0 + \text{Re}(\omega_{//}/pc) \quad (109)$$

The initially monochromatic beam starts wiggling far away from the bucket centers (Fig. 24).

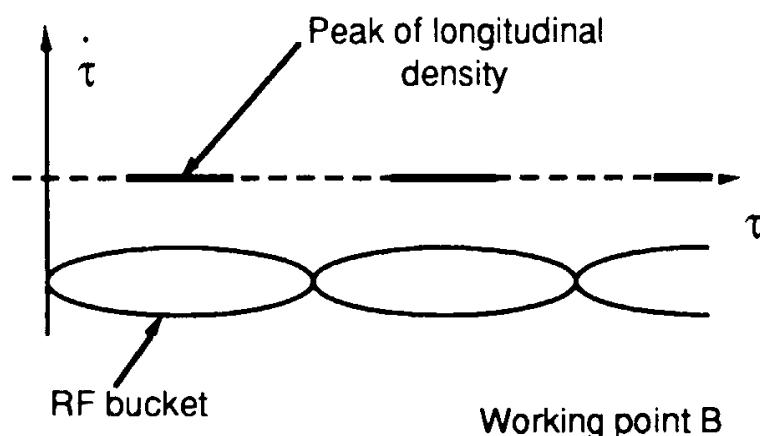


Fig. 24

As already indicated, the resistive part contributes directly to the imaginary part of the frequency shift.

$$\left(\frac{\omega_{\parallel/\text{pc}}}{p\omega_0}\right)^2 = -j\Lambda_{\parallel/\text{pc}} \frac{Z_{\parallel/\text{(p)}}}{p} \quad (110)$$

It is interesting to analyse the solution of the above equation by drawing the curves

$$\text{Re}\left(\frac{Z_{\parallel/\text{(p)}}}{p}\right) \propto \frac{\text{Im}(\omega_{\parallel/\text{pc}})}{p^2} \quad (111)$$

corresponding to a fixed growth rate in the frequency domain.

One can easily conclude that the resistive wall is far less harmful to the beam than the broad-band resistance peaked at high frequencies around the pipe cut-off frequency. In general the instability is of the microwave type with hundreds or thousands of mini-bunches around the circumference.

The conclusion of this section is that a cool beam with a very small momentum spread is unstable with respect to longitudinal instability. This is a problem to be faced by the new generation of cooler rings making use of either stochastic or electron cooling. There exists a threshold at which the instability growth rate is equal to the cooling rate. Under this threshold, the momentum spread can no longer be reduced.

The case of a beam with momentum spread is dealt with in the next section.

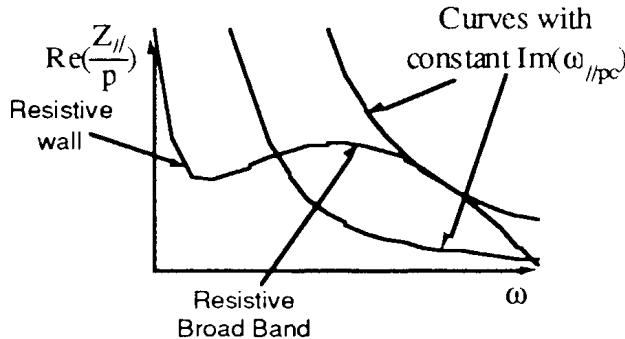


Fig. 25

## 11. COASTING BEAM WITH MOMENTUM SPREAD

It has been shown that in a ring, a very cool beam is subject to microwave instabilities which lead to a beam blow up in momentum. In this section, we will comment on the solution of the Dispersion Relation when a momentum spread is assumed. It will be shown that a sufficient momentum spread in the beam can cure the instability. Different realistic momentum distributions can be assumed. We will treat the case of a parabolic spread:

$$g_0(\dot{\tau}) = \frac{3\omega_0}{8\pi\dot{\tau}_L} \left\{ 1 - \left( \frac{\dot{\tau}^2}{\dot{\tau}_L^2} \right) \right\} \quad \text{for } |\dot{\tau}| \leq \dot{\tau}_L$$

and  $\int_{-\dot{\tau}_L}^{\dot{\tau}_L} g_0(\dot{\tau}) d\dot{\tau} = \frac{\omega_0}{2\pi}$  as required

(112)

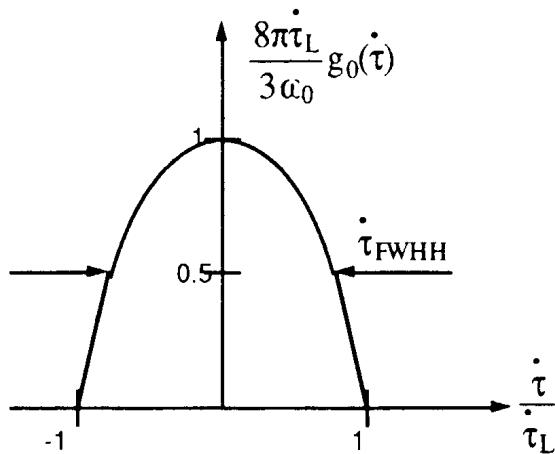


Fig. 26

The important quantity is the spread. Different distribution models with the same Full Width at Half Height spread behave almost identically. Only minor differences can be expected.

With the present model,

$$\dot{\tau}_{FWHH} = \sqrt{2}\dot{\tau}_L = \sqrt{2}|\eta| \left( \frac{dp}{p_{//}} \right)_L = |\eta| \left( \frac{dp}{p_{//}} \right)_{FWHH} \quad (113)$$

the Dispersion Relation can be split into real and imaginary parts:

$$\Lambda_{//c} \operatorname{Re}\left(\frac{Z_{//}}{p}\right) = \operatorname{Im}(J^{-1}) \text{ and } \Lambda_{//c} \operatorname{Im}\left(\frac{Z_{//}}{p}\right) = -\operatorname{Re}(J^{-1}) \quad (114)$$

by making use of the following notations:

$$\Lambda_{//c} = \frac{3 \frac{q}{A}}{2\pi n \frac{m_0 c^2}{e} \gamma_0 \beta_0^2 \left( \frac{dp}{p_{//}} \right)^2 \text{FWHH}}$$

$$J = \int_{-1}^{+1} \frac{x dx}{x + \frac{\omega_{//} pc}{p \omega_0 \tau_L}}$$
(115)

In the stability diagram, the curves corresponding to a constant value of:

$$U = \text{Im}\left(\frac{\omega_{//} pc}{p \omega_0 \tau_L}\right) = \text{Im}\left(\frac{\sqrt{2} \omega_{//} pc}{p \omega_0 n \left( \frac{dp}{p_{//}} \right) \text{FWHH}}\right)$$
(116)

can be drawn.

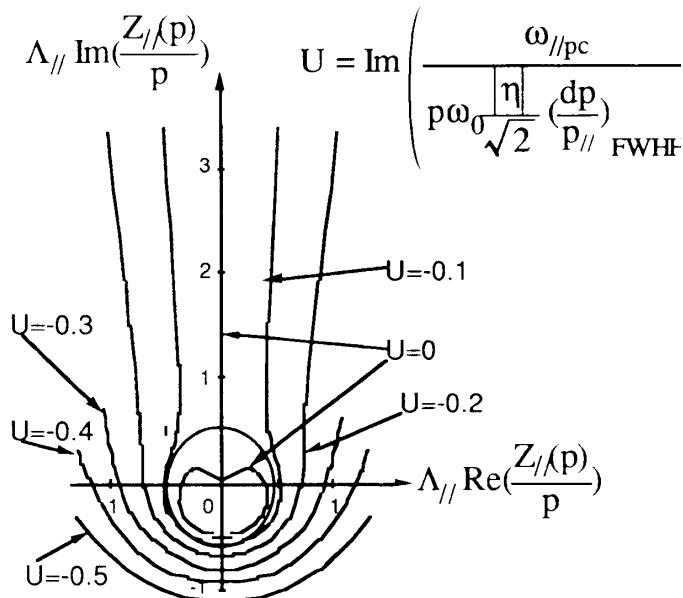


Fig. 27

The stability limit corresponds to the curve  $U = 0$ . It is divided into two parts, the heart shape curve around the origin and the positive part of the vertical axis.

The actual shape of the stability limit depends on the distribution edges. Sharp edge distributions are less stable. A small rounding of the edges makes the contour more circular in the direction of the vertical axis. Distributions with long tails like a gaussian for instance are stable along the lower part of the vertical axis. When one neglects this distribution edge effect, the stability limit can be approximated by a circle.

$$\begin{aligned} \Lambda/c \frac{|Z_{\parallel}/(p)|}{\pi} &\leq 0.5 \quad \text{or} \\ \left| \frac{Z_{\parallel}/(p)}{\pi} \right| &\leq \frac{m_0 c^2}{e} \frac{\gamma_0 \beta_0^2 |\eta|}{\frac{q}{A} I} \left( \frac{dp}{p_{\parallel}} \right)^2_{\text{FWHH}} \end{aligned} \quad (117)$$

The minimum momentum spread required for stability is:

$$\left( \frac{dp}{p_{\parallel}} \right)^2_{\text{FWHH}} \geq \frac{\frac{q}{A} I}{\frac{m_0 c^2}{e} |\eta| \gamma_0 \beta_0^2} \left| \frac{Z_{\parallel}/(p)}{p} \right| \quad (118)$$

## 12. LANDAU DAMPING BY MOMENTUM SPREAD

We can summarize the results of the last two sections as follows:

- 1) A monochromatic beam is unstable and the coherent frequency shift is given by:

$$\left( \frac{\omega_{\parallel}/pc}{p\omega_0} \right)^2 = \frac{-\eta \frac{q}{A} I}{2\pi \frac{m_0 c^2}{e} \gamma_0 \beta_0^2} j \frac{Z_{\parallel}/(p)}{p} \quad (119)$$

- 2) Stability requires a minimum spread in incoherent frequencies

$$\left( \frac{d\omega}{\omega_0} \right)^2_{\text{FWHH}} = \eta^2 \left( \frac{dp}{p_{\parallel}} \right)^2_{\text{FWHH}} \geq \frac{\frac{q}{A} I}{\frac{m_0 c^2}{e} |\eta| \gamma_0 \beta_0^2} \left| \frac{Z_{\parallel}/(p)}{p} \right| \quad (120)$$

It is clear that around the harmonic  $p$  of the revolution frequency the incoherent frequency band is  $p d\omega$  wide. The two formulae can be combined and the result can be expressed as follows: Stability requires the monochromatic beam coherent frequency shift to remain sufficiently inside the incoherent frequency band.

$$\left| \frac{\omega_{\parallel}/pc}{p\omega_0} \right| \leq \sqrt{\frac{2}{\pi}} \left( \frac{d\omega}{\omega_0} \right)_{\text{HWHH}} = \sqrt{\frac{2}{\pi}} \left( \frac{pd\omega}{p\omega_0} \right)_{\text{HWHH}} \quad (121)$$

HWHH stands for Half Width at Half Height

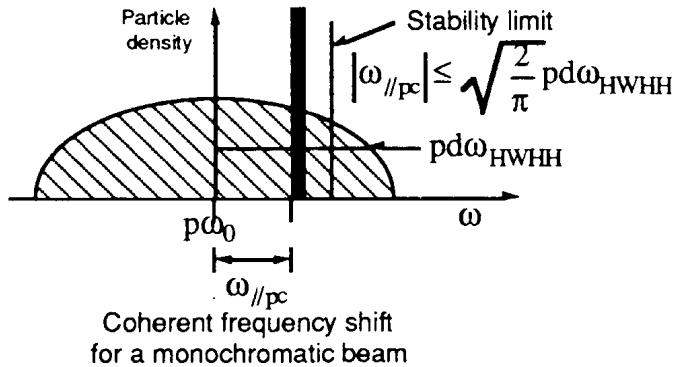


Fig. 28 Coherent frequency shift for a monochromatic beam

The stabilizing effect linked with the incoherent frequency spread is a general mechanism called Landau-Damping. The way it physically acts against the forces which drive the instability is hidden in Vlasov's equation which was used to establish the Dispersion Relation.

Let us try at least to qualitatively explain this Landau-Damping. We will proceed in two steps:

- We first ignore the self force and assume at  $t = 0$  a prebunching of the beam with  $p$  wave lengths around the circumference.
  - i) With a monochromatic beam, all particles have the same revolution frequency; the initial prebunching remains forever and a spectrum analyser indicates an infinitely narrow line at  $p\omega_0$
  - ii) Now, we introduce a spread in momentum and therefore in revolution frequency in the beam. The initial prebunching will disappear.

After a time interval

$$\Delta t = \frac{2\pi}{pd\omega_{FWHH}} \quad (122)$$

two particles which have a difference in revolution frequency equal to  $d\omega_{FWHH}$  and which had started from the same position at  $t = 0$  are now separated by one wavelength. Therefore, after  $\Delta t$  the prebunching has disappeared.  $\Delta t$  will be called the coherence time. Obviously, the larger the spread the shorter the coherence time.

After this introduction, we have all the elements to understand Landau-Damping. From now on, we reintroduce the self field. In the case of a monochromatic beam, the growth time of the instability is:

$$\tau_p = \frac{-1}{\text{Im}(\omega_0/pc)} \quad (123)$$

There are two extreme cases.

In the first case, the incoherent frequency band is so narrow that the coherence time is much longer than the growth time of the monochromatic beam. Then, the beam behaves essentially like a monochromatic beam and blows up.

At the other extreme, the incoherent frequency band is so wide that the coherence time is much shorter than the monochromatic beam growth time. Then, the edges of the distribution

are rapidly out of phase with respect to the driving force. They do not contribute to the coherent signal anymore nor as a consequence to the amplitude of the force. The beam remains stable.

The regime where the beam is just at the stability limit stands between these two extreme cases. One can advance that the threshold is reached when the coherence time is of the order of the growth time.

$$- \text{Im}(\omega_{\parallel}/pc) \approx pd\omega_{\text{FWHH}} \quad (124)$$

For comparison, in our summary of the results we noted:

$$|\omega_{\parallel}/pc| \leq \sqrt{\frac{2}{p}} pd\omega_{\text{FWHH}} \quad (125)$$

The order of magnitude is correct.

### 13. LIMITS OF THE THEORY

Several generations of accelerator physicists have worked on this problem of coherent instabilities. The synthesis of their work presented in the previous sections is a powerful and well established theory. This should not hide two major difficulties.

The first difficulty is that before applying this theory, one has to know the impedance of the ring perfectly. Unfortunately this is very difficult and even though a continuous effort is being made on the subject, the design of vacuum chambers is still largely empirical. Development of tools allowing for a more reliable approach to the design of vacuum chambers is progressing. A certain number of numerical codes solving Maxwell's equations in a structure excited by the passage of a mini bunch exist. They have been successfully tested on several types of structures. An example is shown for the PETRA multicell RF cavities traversed by a short gaussian bunch.

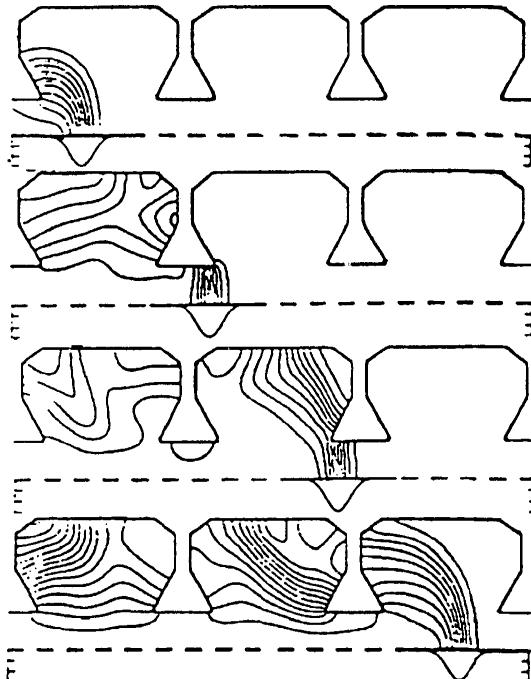
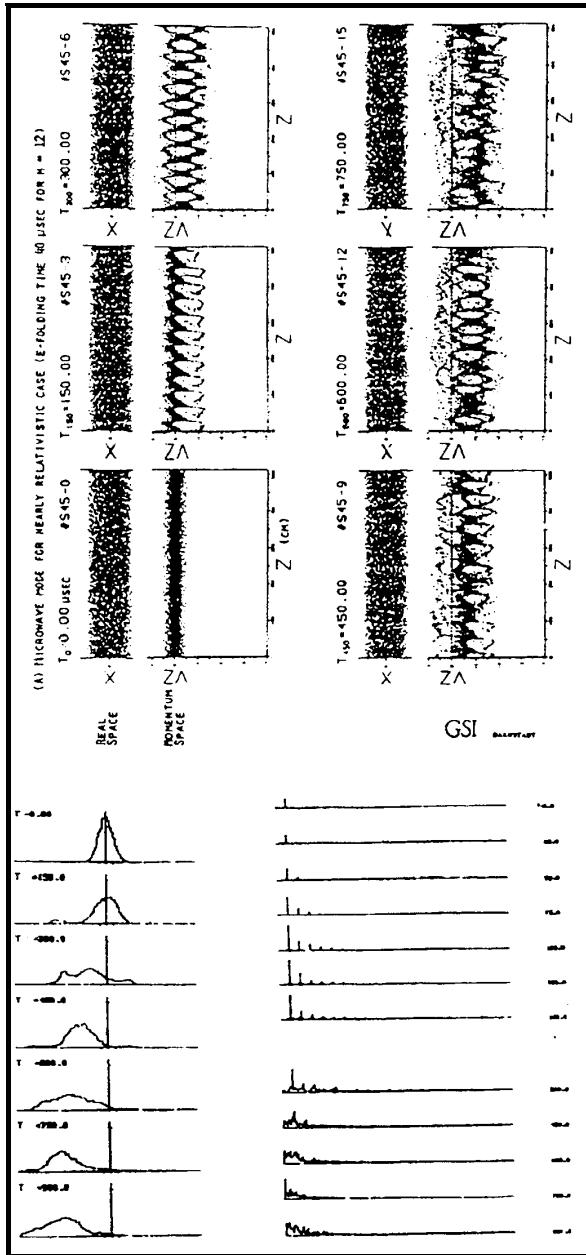


Fig. 29 Fields excited by a Gaussian bunch ( $\sigma = 2$  cm) traversing a PETRA cavity

The second difficulty is that the theory has its own limitations. It is a perturbation theory which uses Vlasov's equation to first order. Therefore, it applies to conditions where the

instability remains very weak. Furthermore, the result drawn from this theory is a single complex number, namely the frequency shift. This is certainly an essential piece of information. However, in the case of unstable conditions, one cannot have access to the time evolution of the beam deterioration. To answer such a question, the only possibility is again to use numerical simulation codes working in the time domain.

An example of simulation is given in the next figure. It shows the development of a microwave instability. The beam-environment coupling assumes a broad band impedance centered around the pipe cut-off frequency. The simulation starts with an initial 10% longitudinal modulation of the particle density at the pipe cut-off frequency corresponding to harmonic 10 in the present case.



Evolution of Momentum Distribution   Evolution of Fourier Spectra

Fig 30 Development of a longitudinal microwave instability

## REFERENCES

- [1] C.E. Nielsen et al., Proc. of Int. Conf. on High Energy Accel. and Instr., CERN (1959) 239
- [2] V.K. Neil, R.J. Briggs, Plasma Physics, 9 (1967) 631
- [3] V.K. Neil, A.M. Sessler, Rev. Sci. Instr. 36 (1965) 429
- [4] B.W. Montague, Single Particle Dynamics, 3rd lecture, Int. School of Particle Accel., Erice (1976)
- [5] B.W. Montague, Single Particle Dynamics, 5th and 6th lectures, Int. School of Particle Accel., Erice (1976)
- [6] H.G. Hereward, Landau Damping, Int. School of Particle Accel., Erice (1976)
- [7] K. Hübner, V.G. Vaccaro, CERN report ISR-TH/70-44 (1970)
- [8] B. Zotter, CERN report ISR-GS/76-11 (1976)
- [9] A.G. Ruggiero, V.G. Vaccaro, CERN report ISR-TH/68-33 (1968)
- [10] E. Keil, W. Schnell, CERN report ISR-TH-RF/69-48 (1969)
- [11] F. Sacherer, Proc. 1973 Part. Accel. Conf., San Francisco, IEEE Trans. Nucl. Sci. Vol. NS-20, n°3, 825
- [12] F. Sacherer, CERN report SI/BR/72-5 (1972)
- [13] A. Hofmann, Single beam collective phenomena-longitudinal, Int. School of Part. Accel., Erice (1976)
- [14] I. Hofmann, Non linear aspects of Landau damping in computer simulation of the microwave instability, Computing in Accelerator Design and Operation, Berlin, Sept. 20-23 1983
- [15] J.L. Laclare, Bunched-beam instabilities, 11th Int. conference on High Energy Accelerators, Geneva (July 1980) 526
- [16] J.L. Laclare, Instabilities in storage rings, Proc. of the Symposium on Accelerator Aspects of H I Fusion, Darmstadt FRG (1982), GSI Report 82-8.

# COASTING BEAM TRANSVERSE COHERENT INSTABILITIES

J.L. Laclare  
ESRF, Grenoble, France

## Abstract

Nowadays, the performance of most machines is limited by Coherent Instabilities. It is one of the most important collective effects which prevents the current from being increased above a certain threshold without the beam quality being spoiled. For the transverse instabilities we will follow a very similar approach to the longitudinal case in the previous chapter. Thus, a large number of basic explanations and comments again apply and will not be repeated. Wherever relevant we will insist on the differences. With regard to the scope of the lecture, the physical mechanisms which will be considered throughout can be applied to any type of machine :

- linear accelerators,
- circular accelerators,
- storage rings

and beam:

- bunched beams
- continuous beams.

## 1 . INTRODUCTION

In the longitudinal plane, the coherent motion is driven by a longitudinal modulation of particle density which creates a longitudinal electric field along the beam axis. When reverting back to the example of the round pipe, this self field is associated with a return or image current  $I_w$  flowing downstream and uniformly distributed on the inner side of the chamber wall.

In the transverse case, we will arbitrarily choose the X plane, the perturbation consists of a slight initial transverse displacement of the beam. Due to the focusing of the external guide field, the beam then oscillates from side to side when progressing along the machine.

The first remark is that the total wall current  $I_w$  has the same magnitude as before but is no longer uniformly distributed in the wall. When comparing the new situation with the unperturbed one, the beam sees a differential current  $\pm \delta I_w$  which flows in the opposite direction on either side of the pipe.

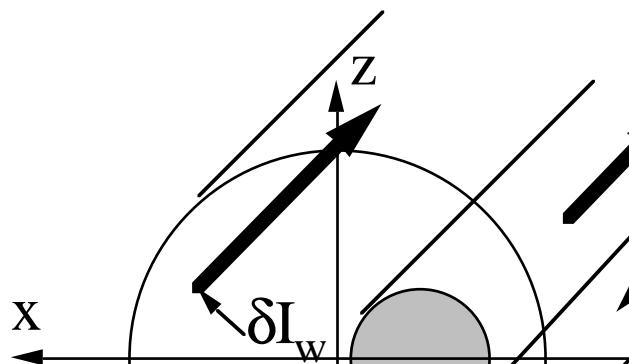


Fig. 1

This requires a longitudinal electric field, which varies linearly in strength across the

aperture, and also a vertical and constant dipole magnetic field. This magnetic field acts back on the beam.

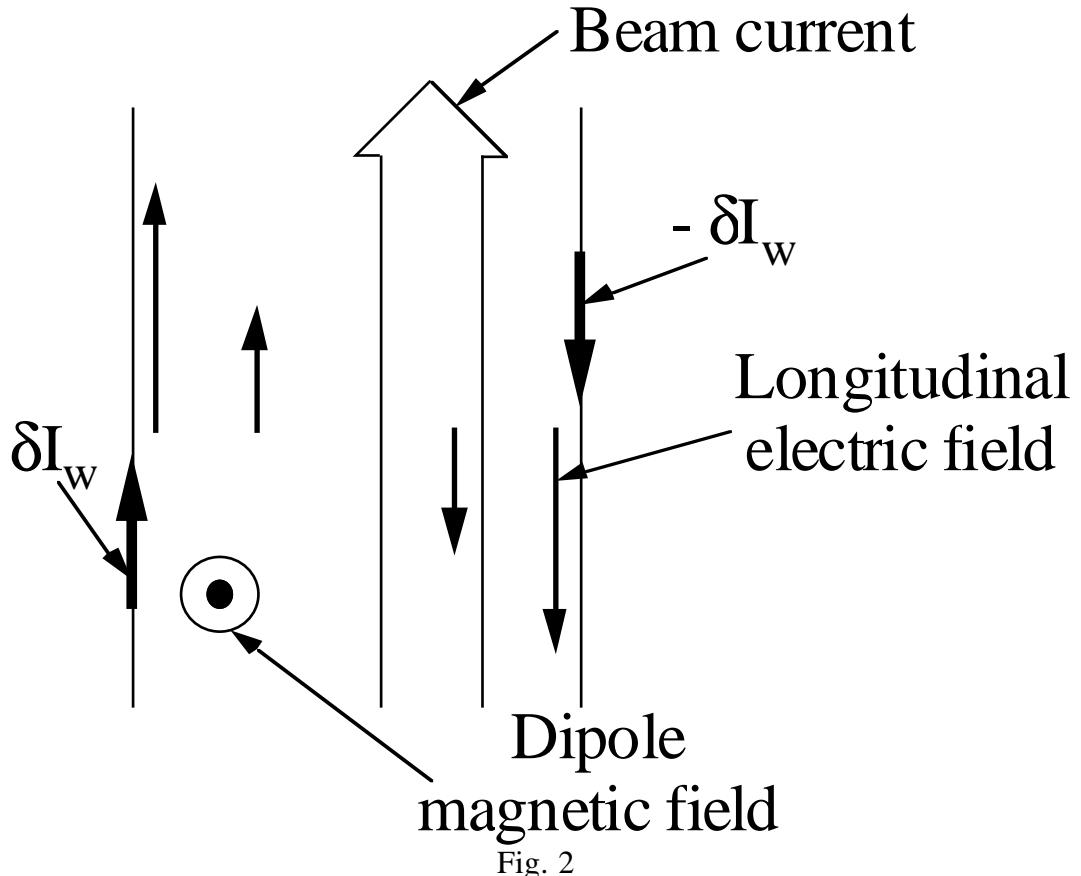


Fig. 2

This magnetic field can either:

- increase the initial displacement thus leading to an instability, or,
- reduce the initial displacement and therefore stabilize the beam.

## 2. SINGLE-PARTICLE TRANSVERSE MOTION

The transverse unperturbed motion is described in the following form:

$$\ddot{x} + \dot{\phi}^2 x = 0 \quad (1)$$

where  $\dot{\phi} = \omega_\beta = Q_x \omega$  is the betatron advance per second,  $Q_x$  the particle horizontal tune, and  $\omega$  the particle revolution frequency.

The transverse position will be expressed as:

$$x(t) = \hat{x} \cos \phi(t) = \hat{x} \cos \left[ \int_0^t \dot{\phi} dt + \phi_0 \right] \quad (2)$$

The couple  $\{\phi, \hat{x}\}$  instead of  $\{x, \dot{x}\}$  will be used for the description of the transverse

motion in canonical variables.

$$\dot{\phi} = \frac{d\phi}{dt} \text{ and } \hat{x} = \left[ x^2 + \left( \frac{\dot{x}}{\dot{\phi}} \right)^2 \right]^{\frac{1}{2}} \quad (3)$$

are two invariants of the unperturbed motion.

Both parameters  $Q_X$  and  $\omega$  depend on momentum  $p_{//}$  and transverse betatron amplitudes  $\hat{x}$  and  $\hat{z}$ .

In the following we will restrict ourselves to:

$$\begin{aligned} \omega &= \omega(p_{//}) \\ \text{and} \\ Q_X &= Q_X(p_{//}, \hat{x}) \end{aligned} \quad (4)$$

Concerning the dependence on momentum, we introduce the following complementary definitions.

$$\xi = \frac{p_{//}}{Q_X} \frac{dQ_X}{dp_{//}} \text{ and } \omega_\xi = Q_{X0}\omega_0 \frac{\xi}{\eta} \quad (5)$$

where  $\xi$  is called the machine chromaticity.

We develop  $\Phi$  with respect to the reference

$$p_{//0}, Q_{X0} .$$

We assume an unperturbed longitudinal motion:

$$\tau = \tau_0 + \dot{\tau}t \text{ with constant } \dot{\tau}$$

and obtain

$$\dot{\phi} = Q_X\omega = Q_{X0}\omega_0(1-\dot{\tau}) + \omega_\xi\dot{\tau} + \dot{\phi}(\hat{x}) \quad (6)$$

$$\phi = \int_0^t \dot{\phi} dt + \phi_0 = Q_{X0}\omega_0(t-\tau) + \omega_\xi\tau + \dot{\phi}(\hat{x})t + \phi_0 \quad (7)$$

Later on, we will perturb the motion by applying the transverse beam self field in the right-hand side of the equation of motion.

$$\ddot{x} + \dot{\phi}^2 x = \frac{e}{m_0 \gamma_0} [\vec{E} + \vec{v} \times \vec{B}]_\perp(t, \theta = \omega_0(t-\tau)) \quad (8)$$

In the  $\{\phi, \hat{x}\}$  plane, the quantity of interest will be

$$\dot{\hat{x}} = \frac{d}{dt} \left[ x^2 + \left( \frac{\dot{x}}{\dot{\phi}} \right)^2 \right]^{\frac{1}{2}} = - \frac{\sin \phi}{\dot{\phi}} (\ddot{x} + \dot{\phi}^2 x) \quad (9)$$

Therefore, the perturbed motion will be studied by means of:

$$\dot{\hat{x}} = - \frac{\sin \phi}{\dot{\phi}} \frac{e}{m_0 \gamma_0} [\vec{E} + \vec{v} \times \vec{B}]_{\perp}(t, \theta = \omega_0(t-\tau)) \quad (10)$$

### 3. TRANSVERSE SIGNAL OF A SINGLE PARTICLE

The transverse PU electrodes deliver a signal proportional to the local beam centre-of-mass position and to the current.

$$S_{\perp} = S_{//} x_{CM} \quad (11)$$

Let us analyse the transverse signal of a single particle.

$$S_{//} = s_{//}(t, \theta) \quad (12)$$

The longitudinal signal consists of a series of periodic Dirac impulses with amplitude  $e$  and period

$$T = \frac{2\pi}{\omega} \quad (13)$$

The transverse amplitude is simply:

$$x = \hat{x} \cos \phi(t) \quad (14)$$

therefore,

$$s_{\perp}(t, \theta) = e \hat{x} \sum_{k=-\infty}^{k=+\infty} \delta(t-\tau - \frac{\theta + 2k\pi}{\omega}) \cos \phi(t) \quad (15)$$

The signal display is sketched below.

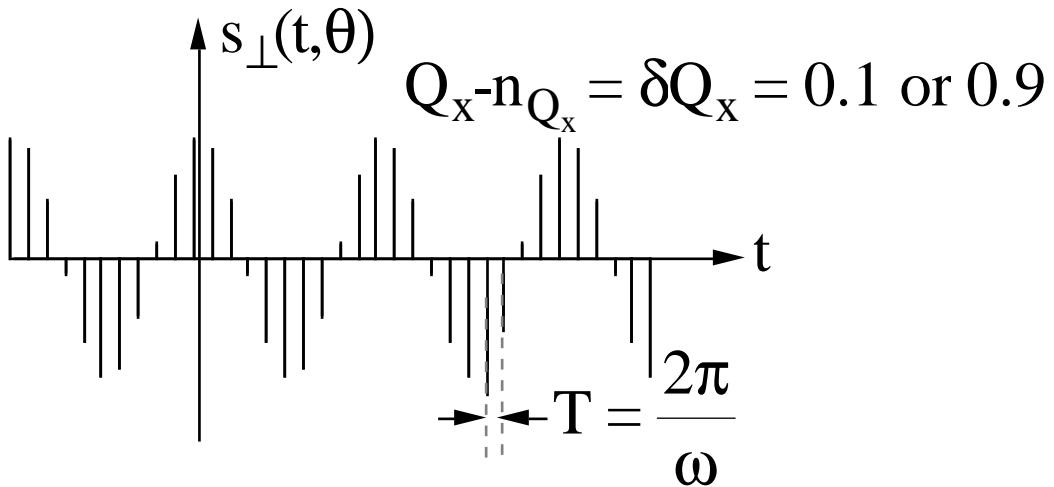


Fig. 3

The wave number is written as:

$$Q_x = n_{Q_x} + \delta Q_x \text{ with } n_{Q_x} \text{ the integer part}$$

$$\text{and } 0 < \delta Q_x < 1 \text{ the non integer part} \quad (16)$$

Beam sampling at the position of the PU turn after turn makes it impossible to identify (modulo 1) the integer part  $n_{Q_x}$  of the wave number. As a consequence, the same signal is obtained for

$$\delta Q_x \text{ and } 1 - \delta Q_x$$

After development of the series of Dirac's, the signal expression becomes:

$$s_{\perp}(t, \theta) = \frac{e\omega_0 \hat{x}}{4\pi} \sum_{p=-\infty}^{p=+\infty} \exp[-j(p\theta + \phi_0)] [\exp j\omega_p^+ t + \exp j\omega_p^- t] \quad (17)$$

with two series of frequencies:

$$\omega_p^+ = (p + Q_x)\omega$$

$$\omega_p^- = (p - Q_x)\omega$$

Thus, in the spectrum, around every harmonic of the revolution frequency, there are two betatron (upper and lower) sidebands.

It is interesting to look for the incoherent frequency spread induced by a momentum spread and a betatron amplitude spread.

$$\delta\omega_p^+ = [(p + Q_x)\omega_0 - \omega_\xi]\dot{\tau} + \dot{\phi}(\hat{x})$$

$$\delta\omega_p^- = [(p - Q_x)\omega_0 + \omega_\xi]\dot{\tau} - \dot{\phi}(\hat{x}) \quad (18)$$

The spread arising from momentum via the chromaticity varies with  $p$ . On the contrary, the spread due to amplitude spread is constant whatever the value  $p$  of the harmonic of the

revolution frequency.

As an example, for the lower sidebands, the incoherent spread due to momentum vanishes for

$$p = Q_{x0} - \frac{\omega_\xi}{\omega_0} \quad (19)$$

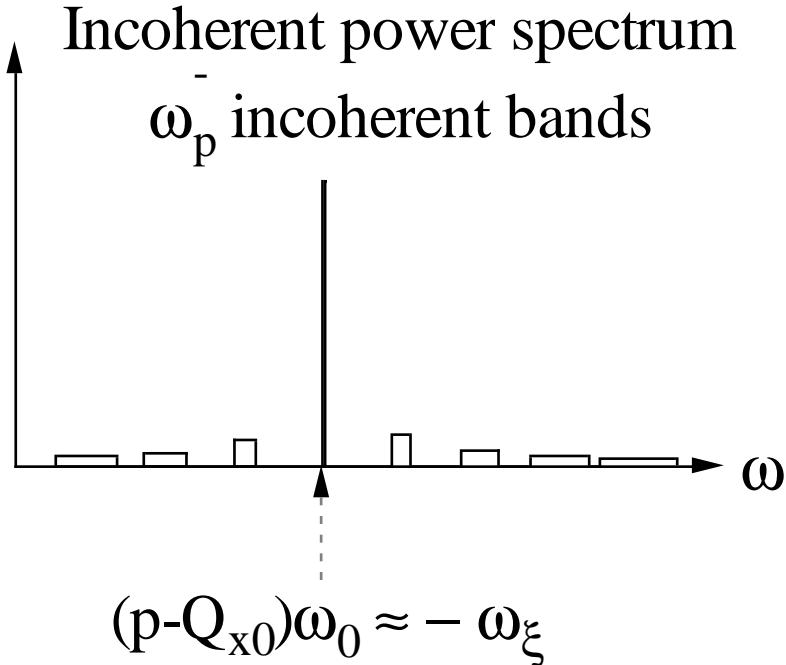


Fig. 4

As seen on the above figure, for this value of  $p$  the incoherent band is very narrow. The same picture would apply for the upper sidebands around  $(p + Q_{x0})\omega_0 \approx \omega_\xi$ .

A priori, these narrow betatron frequency bands represent a potential danger since there will be minimum Landau damping by momentum spread at this frequency. One will have to avoid chromaticity tunes which associate these narrow sidebands with large values of the transverse impedance.

In the following we will arbitrarily privilege the lower sidebands whenever it is necessary to solve the dispersion relation, or when we support our comments by figures. However, lower and upper sidebands are intimately linked and indissociable. We will further develop this fact in the following paragraphs.

#### 4. DISTRIBUTION FUNCTION

To study transverse instabilities, we consider the following distribution:

$$\begin{aligned} \Psi(\varphi, \hat{x}, \tau, \dot{\tau}, t) &= \Psi_0 + \Delta\Psi_p \\ g_0(\dot{\tau})f_0(\hat{x}) + g_p(\dot{\tau})f_p(\varphi, \hat{x}) \exp j(p\omega_0\tau + \omega_{\perp}pc^t) \end{aligned} \quad (20)$$

The first term  $\Psi_0 = g_0(\hat{\tau})f_0(\hat{x})$  is the stationary part normalized as follows:

$$\int_{\hat{\tau}} g_0(\hat{\tau})d\hat{\tau} = \frac{\omega_0}{2\pi} \text{ and } \int_{\hat{x}} f_0(\hat{x})\hat{x}d\hat{x} = \frac{1}{2\pi} \quad (21)$$

The second part  $\Delta\psi_p$  is the perturbation. Due to the  $\tau$  dependence, it consists of a sinusoidal ( $\exp j(p\omega_0\tau)$ ) transverse displacement of the beam with  $p$  wavelengths around the circumference.

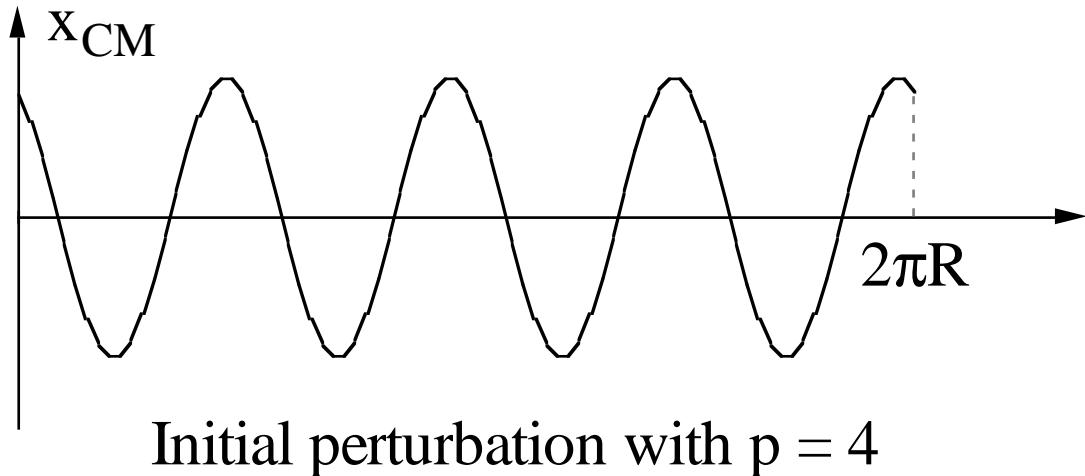


Fig. 5

It is also assumed that this perturbation will move at a coherent frequency  $\omega_{\perp pc}$  apart from  $p\omega_0$ . The frequency offset  $\omega_{\perp pc}$  is a complex number. For a lower sideband, we will write:

$$\omega_{\perp pc} = -Q_{x0}\omega_0 - \Delta Q_{xpc}\omega_0 \quad (22)$$

It is the sum of two quantities:

- the central incoherent  $Q_{x0}\omega_0$  betatron frequency and,
  - the coherent  $\Delta Q_{xpc}\omega_0$  transverse betatron frequency shift.
- The stability is evaluated from the expression

$$\text{Im}(\omega_{\perp pc}) = -\omega_0 \text{ Im}(\Delta Q_{xpc}) \quad (23)$$

## 5 . TOTAL BEAM SIGNAL

In order to measure the total signal induced by the beam, one has to integrate the single particle signal over the distribution.

$$S_{\perp}(t, \theta) = \int_V \Psi(\varphi, \hat{x}, \tau, \dot{\tau}, t) s_{\perp}(t, \theta) dv \quad (24)$$

Let  $\Sigma$  represent the following integral:

$$\Sigma = \int_{\varphi, \hat{x}, \dot{\tau}} g_p(\dot{\tau}) f_p(\varphi, \hat{x}) \hat{x}^2 \cos \varphi d\varphi d\hat{x} d\dot{\tau} \quad (25)$$

Then, the transverse signal can be written:

$$S_{\perp}(t, \theta) = \frac{2\pi I \Sigma}{\omega_0} \exp j[(p\omega_0 + \omega_{\perp pc})t - p\theta] \quad (26)$$

in the time domain and,

$$S_{\perp}(\Omega, \theta) = \frac{2\pi I \Sigma}{\omega_0} \exp (-jp\theta) \delta(\Omega - (p\omega_0 + \omega_{\perp pc})) \quad (27)$$

in the frequency domain. The chosen form of our perturbation leads to a single frequency line which will be used to probe the environment.

## Spectrum analyser

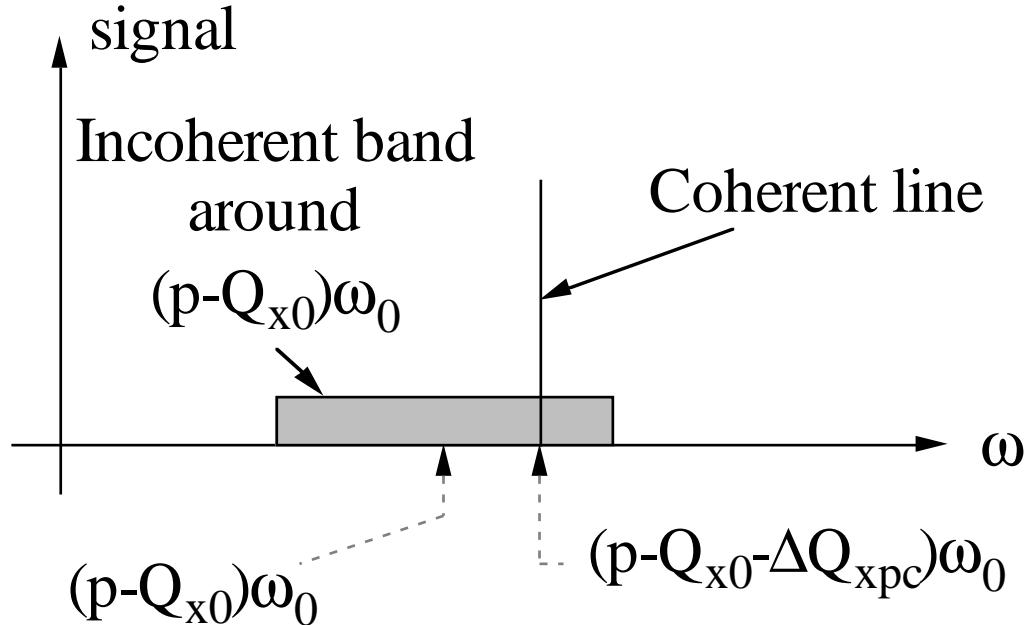


Fig. 6

## 6. DEFINITION OF TRANSVERSE COUPLING IMPEDANCE

Now we introduce the transverse coupling impedance. It relates transverse electromagnetic field and transverse signal at a given frequency as follows:

$$[\vec{E} + \vec{v} \times \vec{B}]_{\perp}(t, \theta) = \frac{-j\beta_0}{2\pi R} Z_{\perp} S_{\perp}(t, \theta) \quad (28)$$

with the units: the signal  $S_{\perp}$  in Am and transverse impedance  $Z_{\perp}$  in  $\Omega m^{-1}$ .

The generalization to any signal is straightforward, it combines spectrum amplitude and impedance over the entire frequency range:

$$[\vec{E} + \vec{v} \times \vec{B}]_{\perp}(t, \theta) = \frac{-j\beta_0}{2\pi R} \int_{\omega=-\infty}^{\omega=+\infty} Z_{\perp}(\omega) S_{\perp}(\omega, \theta) \exp(j\omega t) d\omega \quad (29)$$

In general, one cannot analytically express  $Z_{\perp}(\omega)$ . The solution of Maxwell's equations is never simple. To support the above definition, we just write down the results that one would tediously obtain for a continuous pipe of radius  $b$  through which a circular beam of constant radius  $a$  would travel.

For frequencies below the pipe cut-off frequency:

$$\omega < \omega_{\text{pipe cut-off}} = \frac{c}{b} \quad (30)$$

one could write:

$$[\vec{E} + \vec{v} \times \vec{B}]_{\perp}(t, \theta) = \frac{-j\beta_0}{2\pi R} \left\{ \frac{-jRZ_0}{\beta_0^2 \gamma_0^2} \left( \frac{1}{a^2} - \frac{1}{b^2} \right) \right\} S_{\perp}(t, \theta) \quad (31)$$

provided the pipe wall is perfectly conducting  $\sigma = \infty$ .

The quantity between brackets is the transverse space-charge impedance:

$$Z_{\perp SC} = \frac{-jRZ_0}{\beta_0^2 \gamma_0^2} \left( \frac{1}{a^2} - \frac{1}{b^2} \right) \quad (32)$$

is a pure imaginary quantity and constant in the low frequency domain.

For a non-perfectly-conducting wall,  $\sigma \neq \infty$ , one has to add the resistive-wall contribution. One would find that the transverse and longitudinal resistive-wall impedances are linked by:

$$Z_{\perp RW} = \frac{2c}{b^2} \frac{Z_{\perp}/RW}{\omega} \quad (33)$$

This handy formula is only valid for the resistive wall contribution of a circular pipe. It shows that the curves (impedance diagram)

$$Z_{\perp RW} \text{ and } \frac{Z//RW}{\omega}$$

will be very similar.

For completeness, let us recall that

$$Z//RW = (1+j) \frac{Z_0 \beta_0}{2b} \delta_0 \left( \frac{\omega}{\omega_0} \right)^2 \quad (34)$$

$\delta_0 = \frac{2}{\mu \sigma \omega_0}$   
in which  $\delta_0$  is the skin depth at frequency  $\omega_0$ .

## 7. TRANSVERSE COUPLING IMPEDANCE $Z//(w)$ OF AN ACCELERATOR RING

In qualitative terms, observations made on several machines agree with the following description for the transverse coupling impedance of an accelerator ring. There are four major components. The space-charge impedance was already introduced in a previous section under crude assumptions. Below the pipe cut-off frequency its expression is:

$$Z_{\perp SC} = \frac{-jRZ_0}{\beta_0^2 \gamma_0^2} \left( \frac{1}{a^2} - \frac{1}{b^2} \right) \quad (35)$$

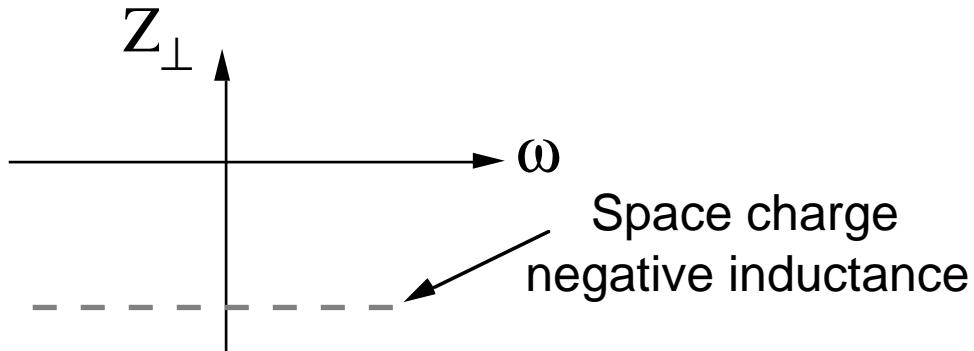


Fig. 7

It is a negative inductance large for slow particles in a pipe of small cross section.

The resistive-wall impedance was also previously mentioned. It will be shown later on that this component of the impedance is the main source of transverse instabilities. Provided the skin depth  $\delta$

$$\delta^2 = \frac{2}{\mu \sigma \omega} \quad (36)$$

at the considered frequency and  $\omega$  is thinner than the wall thickness  $\delta_w$ , then the resistive wall impedance can be expressed as:

$$Z_{\perp RW} = (1+j) \frac{RZ_0}{b^3} \delta_0 \left( \frac{\omega_0}{\omega} \right)^2 \quad (37)$$

At lower frequencies,

$$Z_{\perp RW} = \frac{2cR}{b^3 \sigma \delta_w \omega} \quad (38)$$

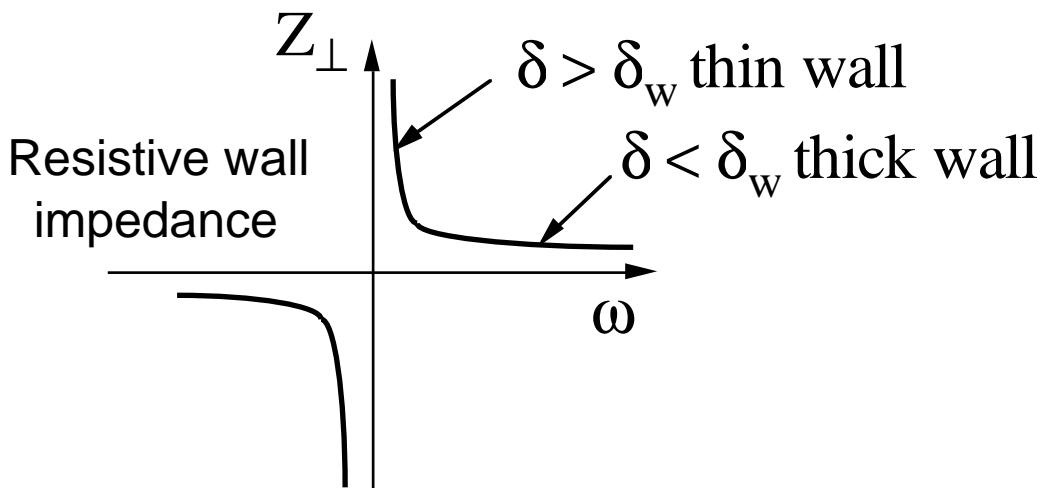


Fig. 8

Due to the  $b^3$  dependence, machines with vacuum chambers of small cross section present a large impedance to the beam.

The third contribution to the machine impedance corresponds to high-Q resonators. As was the case for the longitudinal impedance, the main sources of such resonant objects are RF cavities.

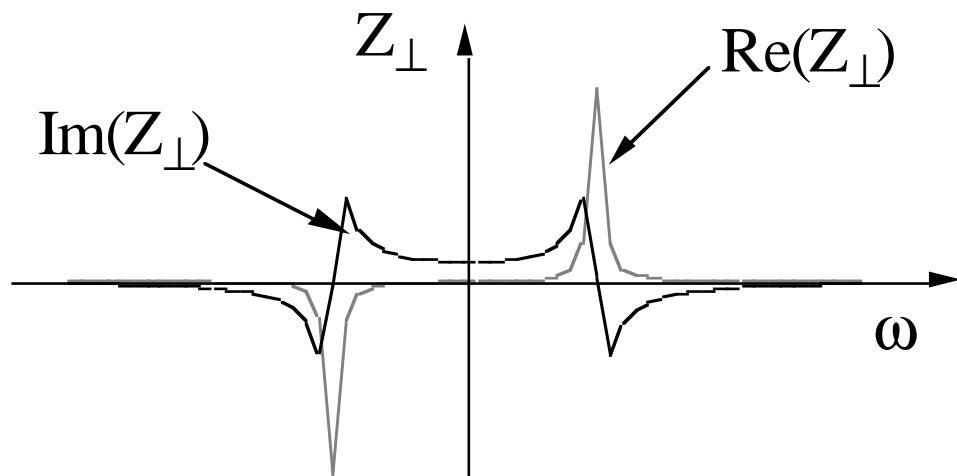


Fig. 9

With regard to the broad-band transverse impedance, up till now we have assumed a perfectly smooth circular pipe with neither cross section changes, bellows, nor flanges, etc. With a view to seeing the difference between an actual chamber and the circular pipe by means of a simple impedance model, it is common practice to introduce a transverse broad-band resonator.

A priori, there is no physical reason to believe that the parameters of this transverse broad-band resonator are linked to the parameters of the longitudinal broad-band resonator. However, measurements made on existing machines show that in the low frequency range the relation

$$Z_{\perp BB} = \frac{2c}{b^2} \frac{Z_{//BB}}{\omega} = \frac{2R}{\beta_0 b^2} \frac{Z_{//BB}}{p} \quad (39)$$

strictly valid for the resistive wake of a circular pipe, can lead to correct orders of magnitudes.

If one remembers that a typical range of longitudinal broad-band impedance value is:

$$0.2 \Omega \leq \left| \frac{Z_{//BB}}{p} \right| \leq 50 \Omega \quad (40)$$

then it can be concluded that the transverse broad-band impedance is in the  $M\Omega/m$  range.

The scaling factor puts large machines with small vacuum chamber cross-section at a disadvantage. However, it is also easier to achieve the lowest values of  $\left| \frac{Z_{//BB}}{p} \right|$  in large machines.

The use of a transverse broad-band impedance model to roughly simulate the effect of abrupt variations of the vacuum chamber, bellows, flanges, etc. leads to the introduction of:

- a positive inductance at low frequencies largely counterbalanced by the space-charge negative inductance for low energy machines,
- a resistive contribution around the pipe cut-off frequency,
- a capacitance at higher frequencies.

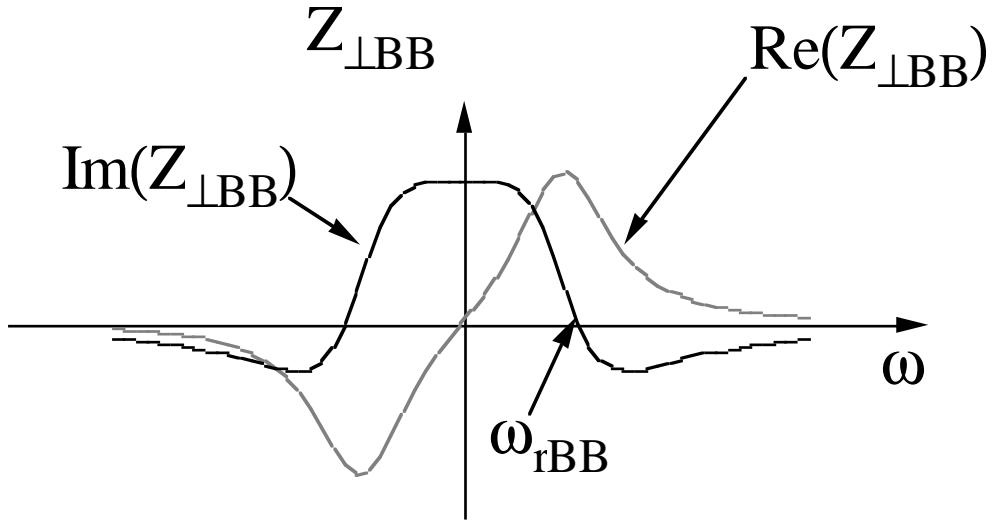


Fig. 10

At this stage it is necessary to point out an essential difference between transverse and longitudinal cases. The longitudinal motion is slow when compared to the revolution period. It takes many turns to perform a complete synchrotron oscillation and one can, in most cases, ignore the fact that some of the sources of impedance (RF cavities for instance) are localized.

In the transverse case, this approximation is not valid. A particle performs  $Q_{x0}$  oscillations per turn and the modulation of the amplitude function  $\beta_x$  cannot be ignored. The fact that particles are more (less) sensitive to impedance sources localised in high (low)  $\beta_x$  has to be taken into account. We can continue to assume a smooth machine with uniformly distributed focusing and impedance provided we introduce a kind of effective impedance to which localized objects contribute as follows:

$$Z_{\perp \text{effective}} = \frac{Q_{x0}}{R} (\beta_x Z_{\perp})_{\text{local}} \quad (41)$$

As an example, the narrow transverse modes of RF cavities would contribute less if cavities were installed in low  $\beta_x$  straight sections.

## 8 . DISPERSION RELATION FOR COHERENT MOTION

In this paragraph, we will gather all the results of previous paragraphs and then apply Vlasov's equation to find out consistent solutions for coherent motion. We will obtain a general dispersion relation. The solution of this equation will give the coherent frequency at which the perturbation oscillates.

$$\omega_{\perp \text{pc}} = -Q_{x \text{pc}} \omega_0 \quad (42)$$

The sign of the imaginary part will tell us whether coherent motion is stable or unstable. Hereunder, the successive steps leading to the dispersion relation are summarized.

Distribution

$$\Psi(\varphi, \hat{x}, \tau, \dot{\tau}, t) = \Psi_0 + \Delta\Psi_p \\ g_0(\dot{\tau})f_0(\hat{x}) + g_p(\dot{\tau})f_p(\varphi, \hat{x})\exp j(p\omega_0\tau + \omega_{\perp}pc^t) \quad (43)$$

Notation

$$\Sigma = \int_{\varphi, \hat{x}, \tau} g_p(\dot{\tau})f_p(\varphi, \hat{x})\hat{x}^2 \cos\varphi d\varphi d\hat{x} d\dot{\tau} \quad (44)$$

Transverse signal

$$S_{\perp}(t, \theta) = \frac{2\pi\Sigma}{\omega_0} \exp j[(p\omega_0 + \omega_{\perp}pc)t - p\theta] \quad (45)$$

Electromagnetic field

$$[\vec{E} + \vec{v} \times \vec{B}]_{\perp} S_{\perp}(t, \theta) = \frac{-j\beta_0}{2\pi R} Z_{\perp}(p) S_{\perp}(t, \theta) \quad (46)$$

Differential equation for single-particle motion

$$\dot{\hat{x}} = - \frac{\sin \varphi}{\dot{\varphi}} \frac{e}{m_0 \gamma_0} [\vec{E} + \vec{v} \times \vec{B}]_{\perp}(t, \theta = \omega_0(t - \tau)) \quad (47)$$

Vlasov's equation

$$\begin{aligned} \frac{d\Psi}{dt} = 0 &= \frac{\partial\Psi}{\partial t} + \frac{\partial\Psi}{\partial\varphi} \dot{\varphi} + \frac{\partial\Psi}{\partial\hat{x}} \dot{\hat{x}} + \frac{\partial\Psi}{\partial\tau} \dot{\tau} + \frac{\partial\Psi}{\partial\dot{\tau}} \ddot{\tau} \\ \frac{\partial\Psi}{\partial t} &= j\omega_{\perp}pc g_p(\dot{\tau})f_p(\varphi, \hat{x})\exp j(*) \\ \frac{\partial\Psi}{\partial\varphi} \dot{\varphi} &= \dot{\varphi} g_p(\dot{\tau}) \frac{\partial f_p(\varphi, \hat{x})}{\partial\varphi} \exp j(*) \\ \frac{\partial\Psi}{\partial\hat{x}} \dot{\hat{x}} &= g_0(\dot{\tau}) \frac{\partial f_0(\hat{x})}{\partial\hat{x}} \frac{\sin\varphi}{\dot{\varphi}} \frac{j\Sigma Z_{\perp}(p)}{\frac{m_0 \gamma_0 c}{e}} \exp j(*) \\ &\quad + \text{neglected second order term} \end{aligned} \quad (48)$$

$$\frac{\partial\Psi}{\partial t} \dot{\tau} = jp\omega_0\tau g_p(\dot{\tau})f_p(\varphi, \hat{x})\exp j(*) \quad (49)$$

$\ddot{\tau} = 0$  no longitudinal electromagnetic field

Intermediate equation

$$\begin{aligned}
& j(p\omega_0\tau + \omega_{\perp pc}) g_p f_p + \phi g_p \frac{\partial f_p}{\partial \phi} \\
& = - g_0 \frac{\partial f_0}{\partial \hat{x}} \frac{\sin \phi}{\phi} \frac{I \Sigma j Z_{\perp}(p)}{\frac{m_0 \gamma_0 c}{e}}
\end{aligned} \tag{50}$$

In the perturbation  $g_p f_p$  we separate the  $\Phi$  dependance on one hand and the  $\hat{x}$  and  $\tau$  dependences on the other. In view of this, we introduce an intermediate function:  $h(\hat{x}, \tau)$  and write:

$$g_p(\tau) f_p(\phi, \hat{x}) = h(\hat{x}, \tau) \left( \cos \phi - j \frac{\sin \phi}{\phi} (p\omega_0\tau + \omega_{\perp pc}) \right) \tag{51}$$

Vlasov's equation reduces to:

$$h(\hat{x}, \tau) = - \frac{I c \Sigma j Z_{\perp}(p)}{\frac{m_0 \gamma_0 c^2}{e} (p\omega_0\tau + \omega_{\perp pc})^2 - \phi^2} \frac{g_0 \frac{\partial f_0}{\partial \hat{x}}}{\tag{52}}$$

However, coming back to the original definition of  $\Sigma$ , one can also write:

$$\Sigma = \pi \int_{\hat{x}, \tau} h(\hat{x}, \tau) \hat{x}^2 d\hat{x} d\tau \tag{53}$$

In the previous expression of Vlasov's equation, after multiplication of both sides by  $\pi \hat{x}^2$  and integration with respect to  $\hat{x}$  and  $\tau$  one finally gets the dispersion relation:

$$1 = \frac{-\pi(q/A) I c}{(\frac{m_0 c^2}{e}) \gamma_0} j Z_{\perp}(p) \int_{\hat{x}, \tau} \frac{g_0(\tau) \frac{\partial f_0(\hat{x})}{\partial \hat{x}} \hat{x}^2}{(p\omega_0\tau + \omega_{\perp pc})^2 - \phi^2} d\tau d\hat{x} \tag{54}$$

In the next three paragraphs examples of solutions of the dispersion relation are given.

## 9. BEAM WITHOUT TUNE SPREAD

First, we assume a very cool beam in the longitudinal transverse plane: no momentum spread. In addition the wave number does not depend on the transverse betatron amplitude. In mathematical terms:

$$g_0(\tau) = \frac{\omega_0}{2\pi} \delta(\tau) \quad (55)$$

and

$$\dot{\phi} = Q_{x0}\omega_0(1-\tau) + \omega_\xi\tau + \dot{\phi}(\hat{x}) \text{ with } \dot{\phi}(\hat{x}) = 0 \quad (56)$$

We use the following relation:

$$\int_{\hat{x}} \frac{\partial f_0(\hat{x})}{\partial \hat{x}} \hat{x}^2 d\hat{x} = -2 \int_{\hat{x}} f_0(\hat{x}) \hat{x} d\hat{x} = -\frac{1}{\pi} \quad (57)$$

and obtain:

$$\begin{aligned} (\omega_{\perp pc}^2 - Q_{x0}^2 \omega_0^2) &= \omega_0^2 (Q_{xpc}^2 - Q_{x0}^2) \\ &= \frac{(\frac{q}{A}) I_c}{2\pi(\frac{m_0 c^2}{e}) \gamma_0} j Z_{\perp}(p) \end{aligned} \quad (58)$$

$Q_{xpc}^2$  is the coherent wave number.

The equation always has two roots (lower and upper sidebands):

$$\omega_{\perp pc}^{\pm} = Q_{\perp pc}^{\pm} \omega_0 \quad (59)$$

For small perturbations they are given by:

$$\operatorname{Re}(\omega_{\perp pc}^{\pm}) = \pm \left\{ Q_{x0} \omega_0 - \frac{(\frac{q}{A}) I_c \operatorname{Im}(Z_{\perp}((p \pm Q_{x0}) \omega_0))}{4\pi Q_{x0} (\frac{m_0 c^2}{e}) \gamma_0} \right\} \quad (60)$$

$$\operatorname{Im}(\omega_{\perp pc}^{\pm}) = \pm \frac{(\frac{q}{A}) I_c}{4\pi Q_{x0} (\frac{m_0 c^2}{e}) \gamma_0} \operatorname{Re}(Z_{\perp}((p \pm Q_{x0}) \omega_0)) \quad (61)$$

If the impedance has a resistive part, the coherent motion is always unstable.

The two series of coherent frequencies corresponding to lower and upper sidebands are solutions of the coherent motion:

$$\omega^+ = p\omega_0 + \omega_{\perp pc}^+ \quad \text{and} \quad \omega^- = p\omega_0 + \omega_{\perp pc}^- \quad (62)$$

Due to the properties of the impedance, for opposite values of the frequency  $\omega$ :

$$\begin{aligned}\operatorname{Re}(Z_{\perp}(\omega)) &= -\operatorname{Re}(Z_{\perp}(-\omega)) \\ \operatorname{Im}(Z_{\perp}(\omega)) &= \operatorname{Im}(Z_{\perp}(-\omega))\end{aligned}\quad (63)$$

both series lead to identical results. For instance one can look for conditions for instability:

$$\operatorname{Im}(\omega_{\perp \text{pc}}) < 0 \quad (64)$$

Unstable motion for the upper sidebands when

$$\omega^+ \approx (p + Q_{x0})\omega_0 \text{ is negative}$$

Unstable motion for the lower sidebands when

$$\omega^- \approx (p - Q_{x0})\omega_0 \text{ is positive}$$

This can be sketched in the impedance diagram:

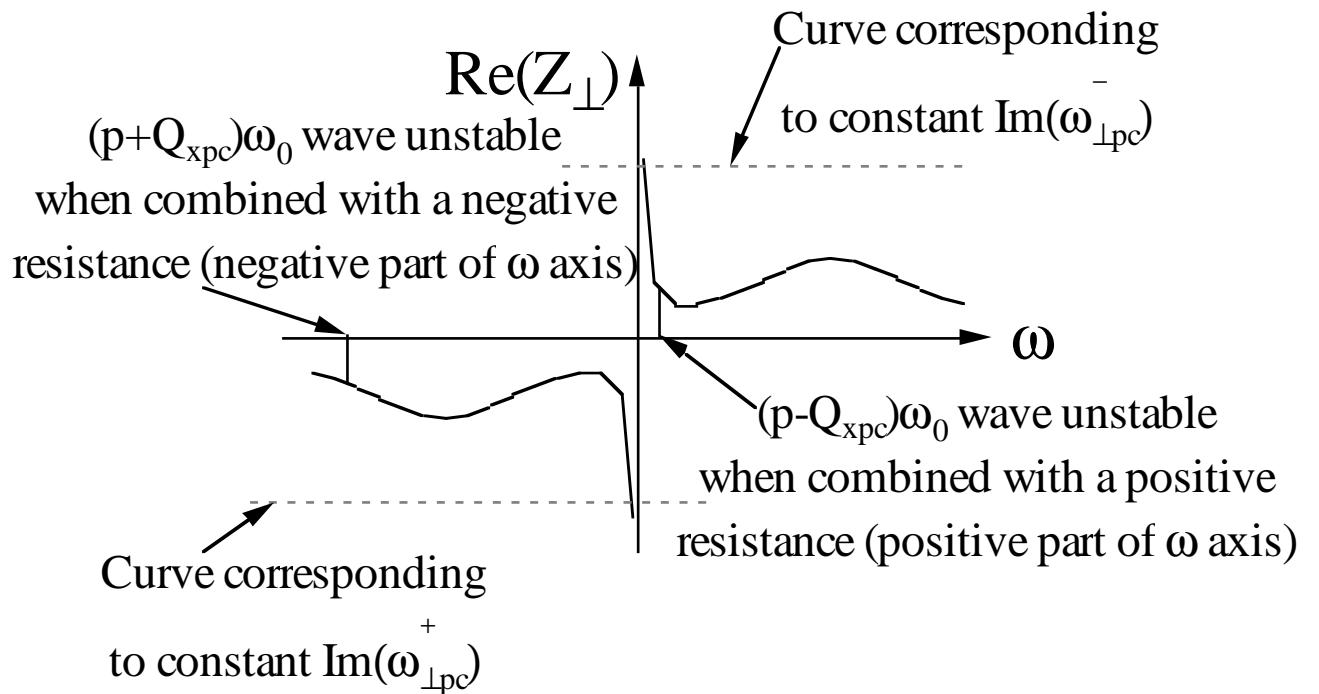


Fig. 11

In this figure, it is obvious that the transverse instability of a coasting beam is essentially a low frequency mechanism. This is because the beam is very sensitive to the low frequency region where the skin resistance tends to be very large, in particular much larger than the broadband resistance.

In the stability diagram with

$$\operatorname{Re}(Z_{\perp}(\omega)) \text{ and } \operatorname{Im}(Z_{\perp}(-\omega))$$

coordinates, the curves corresponding to constant growth rate are vertical lines.

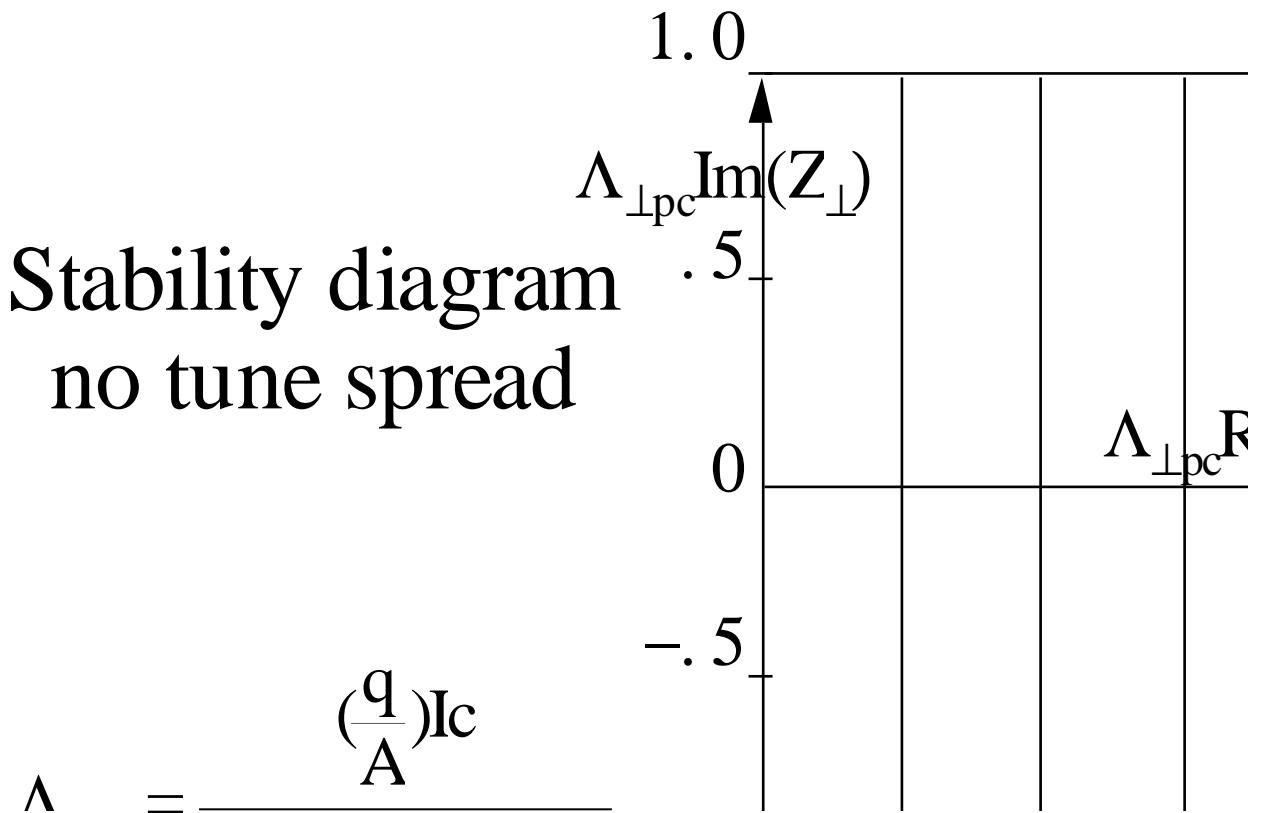


Fig. 12

The beam is always unstable except along the vertical axis. If the impedance were a purely imaginary number (inductance for instance), then the frequency shift would be real and coherent motion would be stable. In this respect, for the transverse plane we do not find the equivalent of the negative mass instability for the longitudinal plane.

When the beam is stable, a tune measurement device (RF knock out for instance) which necessarily detects the coherent motion only, indicates a certain value  $Q_{xpc}$  of the coherent wave number. A priori, our results indicate that  $Q_{xpc}$  is a linear function of beam intensity and one could imagine that the experimental curve  $Q_{xpc}$  versus current would allow the imaginary part of  $Z_{\perp}(p)$  to be measured.

Unfortunately this is not the case. As a matter of fact, the space-charge contribution is not accessible and one will measure

$$\text{Im}(Z_{\perp} - Z_{\perp SC})$$

This specificity of the space-charge component deserves some explanation. Let us assume a perfectly centered intense beam at low energy. The actual wave number of particles oscillating around the beam center, called incoherent wave number, is the result of two quadrupolar fields:

- the focusing of the external guide field
- the space-charge defocusing effect.

This incoherent wave number is the quantity noted  $Q_{x0}$  in this report. In other words:

$$Q_{x0} = Q_x \text{ external guide field} + \Delta Q_x \text{ space charge} \quad (65)$$

Obviously the tune is depressed by space-charge

$$\Delta Q_x \text{ space charge} < 0 \quad (66)$$

However, it must be pointed out that the space-charge field is null at the beam center. Now we rigidly displace the beam center and look for the coherent wave number. The beam center motion is influenced by:

- the focusing from the external guide field,
- the coherent deflecting magnetic field due to the broad-band inductance,
- the coherent deflecting field due to space-charge.

However, the space-charge field is still null at beam center. The coherent wave number logically compensates the incoherent tune depression due to space-charge.

As a conclusion, with a tune measurement device one cannot have access to  $Z_{\perp SC}$  simply because the beam center of mass is not influenced by this field.

A practical remark can be made concerning the choice of the wave number. We have seen that the resistive wall impedance is likely to be the main source of instability. It behaves like

$$\omega^{-0.5}$$

in the thick wall assumption. It is therefore necessary to have the lowest coherent line at a frequency as high as possible

$$\omega \approx (p - Q_{x0})\omega_0$$

In view of this, with a tune 0.1 above an integer, the first coherent frequency line is at

$$\omega \approx 0.9\omega_0$$

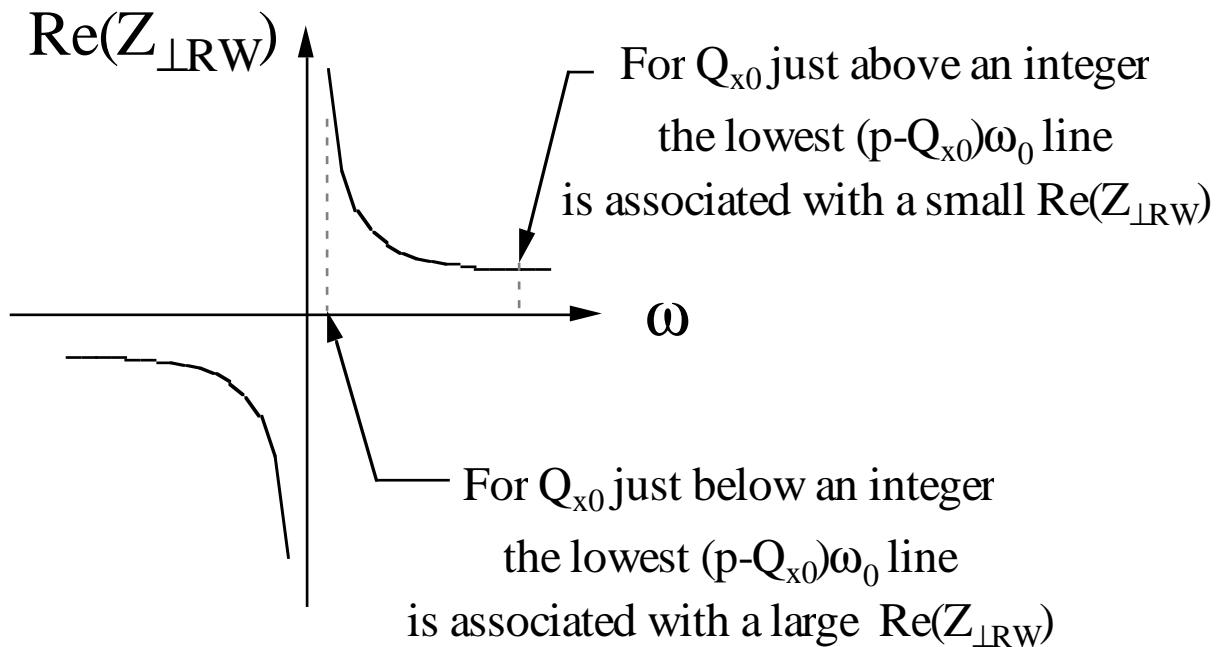


Fig. 13

On the contrary, with a tune of 0.9 (0.1 below the closest integer), this first coherent frequency line is at  $\omega \approx 0.1\omega_0$ . This factor 9 in frequency leads to a factor 3 in the value of corresponding resistive-wall impedance and consequently instability growth rate. Therefore, preference must be given to tunes just above an integer. As shown in this paragraph a beam with no spread in tune is always unstable. A spread in tune can provide the necessary Landau damping.

There are two principal possibilities for providing a tune spread:

- chromaticity via momentum spread, or
- transverse non linearities (tune variation with amplitude).

These two cases are studied independently in the two next paragraphs.

## 10. LANDAU DAMPING BY MOMENTUM SPREAD

We will assume a parabolic stationary distribution in momentum:

$$g_0(\tau) = \frac{3\omega_0}{8\pi\tau_L} \left( 1 - \frac{\tau^2}{\tau_L^2} \right) \quad (67)$$

and solve the dispersion relation.

The denominator of the quantity in the integral can be written as the product of two terms

$$\begin{aligned} & (\omega_{\perp pc} - Q_{x0}\omega_0 + ((p+Q_{x0})\omega_0 - w_\xi)\tau) \\ & (\omega_{\perp pc} + Q_{x0}\omega_0 + ((p-Q_{x0})\omega_0 + w_\xi)\tau) \end{aligned} \quad (68)$$

They are associated with the upper (for the first one) and lower (for the second one) sidebands. We know from the previous paragraph that both waves lead to the same result. We will therefore concentrate on the slow wave (second term above) and look for the solution  $\omega_{\perp pc} \approx -Q_{x0}\omega_0$ . In this case, the first term above can be approximated by  $-2Q_{x0}\omega_0$ .

The then simplified dispersion relation can be written as follows:

$$1 = \frac{-\left(\frac{q}{A}\right)Ic}{2Q_{x0}\omega_0\left(\frac{m_0c^2}{e}\right)\gamma_0} jZ_{\perp}(p) \int_{\tau}^{\infty} \frac{g_0(\tau) d\tau}{\omega_{\perp pc} + Q_{x0}\omega_0 - (p - Q_{x0})\omega_0 - \omega_{\xi}} \quad (69)$$

To simplify the writing we use the following definitions:

$$\Delta\omega_p = \left( (p - Q_{x0})\omega_0 + \omega_{\xi} \right) \left| \eta \left( \frac{\delta p}{p} \right) \right| \quad (70)$$

This quantity represents half the full width band (measured at the foot) of incoherent spread of frequency around the considered lower sideband line  $(p - Q_{x0})\omega_0$ .

$$\Lambda_{\perp pc} = \frac{3\left(\frac{q}{A}\right)Ic}{16\pi Q_{x0}\left(\frac{m_0c^2}{e}\right)\gamma_0\Delta\omega_p} \quad (71)$$

$$x_1 = -\frac{\omega_{\perp pc} + Q_{x0}\omega_0}{\Delta\omega_p} = \frac{(Q_{xpc} - Q_{x0})\omega_0}{\Delta\omega_p} = \frac{\Delta Q_{xpc}\omega_0}{\Delta\omega_p} \quad (72)$$

$x_1$  is the coherent betatron frequency shift normalized to the incoherent spread defined above.

We also use:

$$J_{\perp} = \int_{-1}^{+1} \frac{1-x^2}{x-x_1} dx \quad (73)$$

With these notations the dispersion relation can be finally written:

$$\frac{1}{J_{\perp}} = \Lambda_{\perp pc} j Z_{\perp}(p) \quad (74)$$

The stability diagram with:

$$\Lambda_{\perp pc} \operatorname{Re}(Z_{\perp}(p)) \text{ and } \Lambda_{\perp pc} \operatorname{Im}(Z_{\perp}(p))$$

along the axes is shown in Fig. 14.

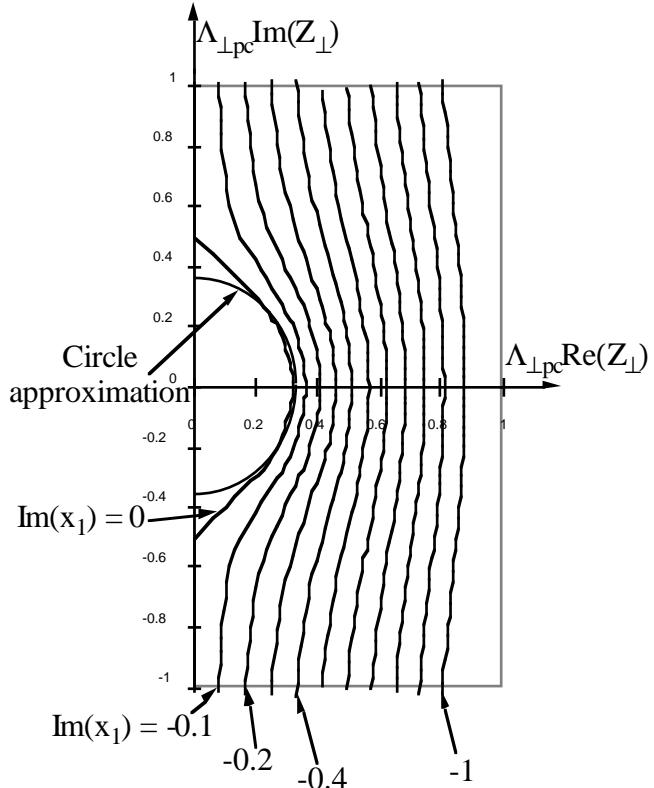


Fig. 14

The imaginary part of the coherent frequency is positive and the movement is damped for small values of the transverse impedance. When compared to the previous example of a monochromatic beam, provided the impedance is reasonably small, the incoherent frequency spread arising from momentum can stabilize the coherent motion. On the contrary, for large  $|Z_{\perp}(p)|$  Landau damping is not strong enough and coherent motion is unstable. The curve  $\operatorname{Im}(x_1) = 0$  defines the stability limit. Its contour can be approximated by a circle:

$$|\Lambda_{\perp pc} Z_{\perp}(p)| < \frac{\sqrt[3]{2}}{4\pi} F \quad (75)$$

with  $F \approx 1$  for the assumed parabolic distribution.

Although the detailed form of the stability limit curve depends on the exact momentum distribution function, very similar results would be obtained for other realistic distribution

functions with the same  $\frac{(\delta p)}{p}$  FWHH. The criterion can be rewritten in terms of the incoherent spread. We will use the FWHH as a reference:

$$\Delta\omega_p \text{FWHH} = \sqrt{2}\Delta\omega_p \quad (76)$$

Then stability requires:

$$\Delta\omega_p \text{FWHH} > \frac{\left(\frac{q}{A}\right)I_c|Z_{\perp}(p)|}{4Q_x 0 \left(\frac{m_0 c^2}{e}\right)\gamma_0} \quad (77)$$

It is interesting to note that the quantity on the right hand side of the above relation can be very simply linked with the coherent betatron frequency shift one would obtain with a monochromatic beam (cf previous paragraph).

$$\Delta\omega_p \text{FWHH} > \pi \omega_0 |Q_{xpc} - Q_{x0}|_{\text{monochromatic beam}} \quad (78)$$

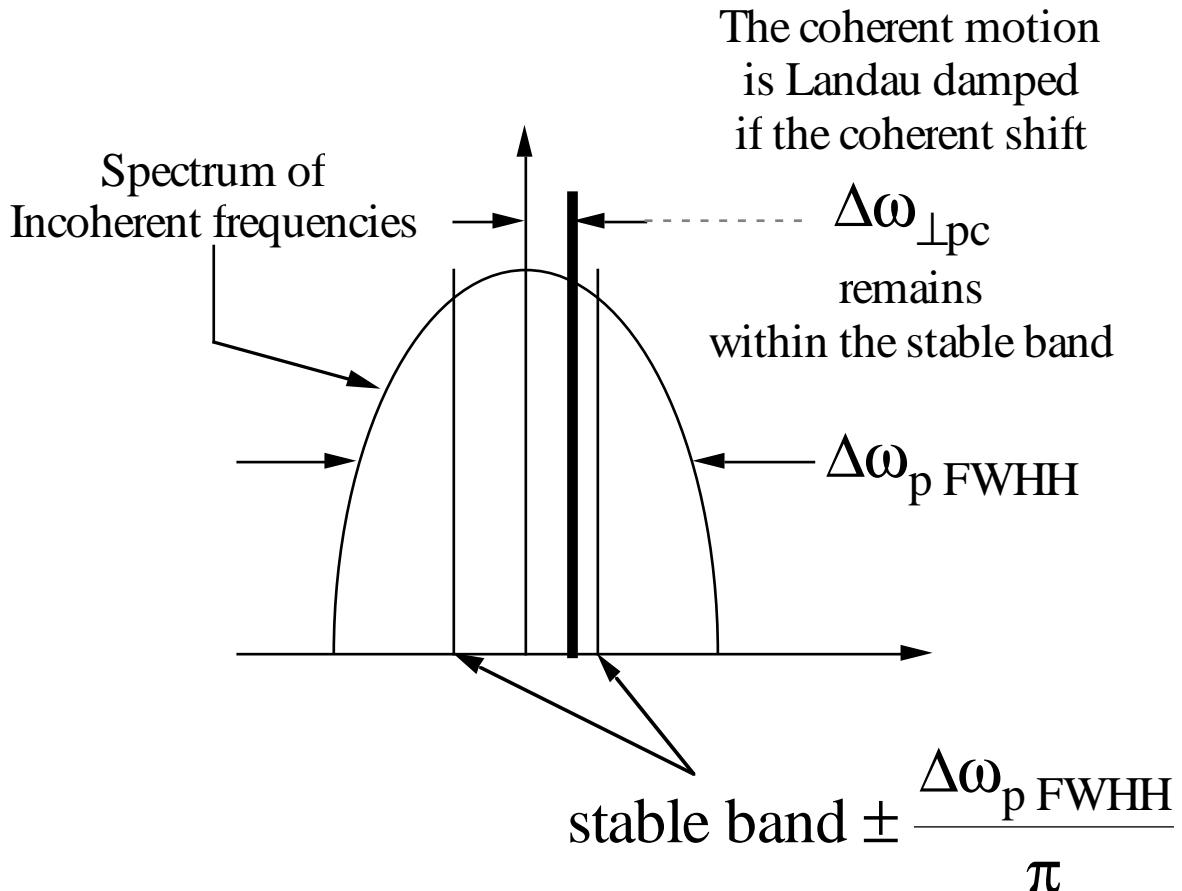


Fig. 15

For standard uncorrected optics,  $\xi \approx -1$  and  $\omega_\xi = \frac{Q_x 0 \omega_0}{\eta}$  are negative above

transition. This corresponds to the worst situation because

$$\Delta\omega_p \text{ FWHH} = ((p - Q_{x0})\omega_0 + \omega\xi) \left| \eta \left( \frac{\delta p}{p} \right) \right|_{\text{FWHH}}$$

$$p \approx Q_{x0} \left( 1 - \frac{\xi}{\eta} \right)$$

vanishes around

To improve the situation, one has to change the sign of the chromaticity by introducing sextupoles in the dispersive sections of the lattice. Correction of chromaticity in both planes is not always simple in particular in large machines. The non-linearities generated by the sextupoles can severely limit the single particle dynamic acceptance of the machine. The optimization of the sextupolar correction scheme is often very challenging. It can largely influence the choice of the basic linear optics.

It must be pointed out that transverse stability of coherent motion is not necessarily the only reason to correct chromaticity. When the uncorrected  $\xi$  is large, the incoherent tune spread due to momentum is also large. In a tune diagram, space between dangerous betatron resonances is always limited.

For machines working below transition, the natural chromaticity is in general positive and has therefore the right sign to always provide some Landau damping. In the impedance diagram, one can draw the line which represents the incoherent frequency band  $\Delta\omega_p \text{ FWHH}$  as a function of  $\omega$ .

It can be seen that the low frequency region is the most dangerous and for two reasons. The resistive wall impedance is large and the frequency band is narrow.

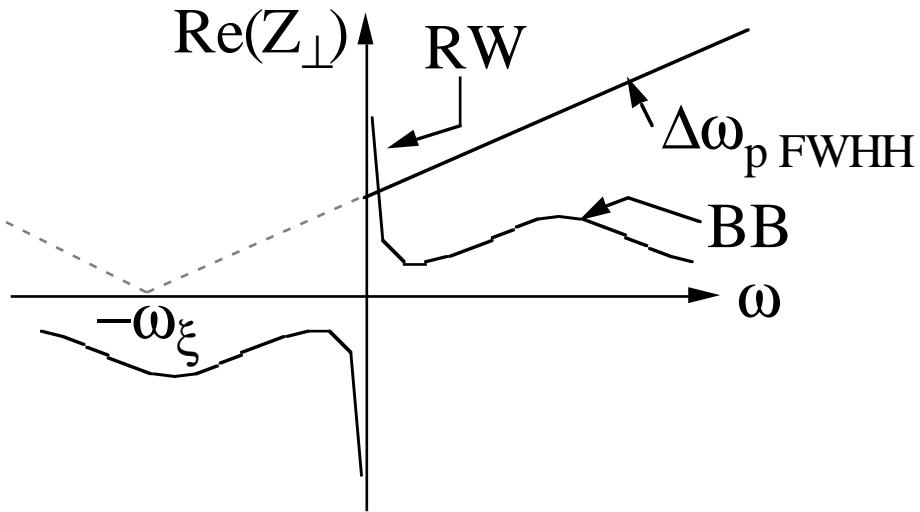


Fig. 16

This is the reason why transverse instability is currently a low-frequency mechanism. If Landau damping is insufficient, coherent motion can also be stabilized by a feedback system. Fortunately, the conception of such a system is easier at low frequency.

## 1. LANDAU DAMPING BY AMPLITUDE DEPENDENT TUNE

Now we separately consider the influence of a tune spread arising from a betatron amplitude spread in the beam. In order to write down the corresponding dispersion relation, we assume a monochromatic beam:

$$g_0(\tau) = \frac{\omega_0}{2\pi} \delta(\tau) \quad (80)$$

and for instance a parabolic distribution of betatron amplitudes:

$$f_0(\hat{x}) = \frac{2}{\pi \hat{x}_L^2} \left( 1 - \left( \frac{\hat{x}}{\hat{x}_L} \right)^2 \right) \text{ for } 0 < \hat{x} < \hat{x}_L$$

and

$$f_0(\hat{x}) = 0 \text{ for } \hat{x} > \hat{x}_L \quad (81)$$

Under these assumptions, the dispersion relation takes the following form:

$$1 = \frac{-\left(\frac{q}{A}\right)I_c}{2\pi Q_x \omega_0 \left(\frac{m_0 c^2}{e}\right) \gamma_0 \Delta Q_L} j Z_{\perp}(p) \int_0^1 \frac{x \, dx}{x - \frac{\Delta Q_x p c}{\Delta Q_L}} \quad (82)$$

The following definition:

$$\Delta Q_L = \frac{\partial Q_x}{\partial \hat{x}^2} \hat{x}_L^2 \quad (83)$$

is used to measure the total incoherent tune spread due to amplitude in the beam. We then note:

$$x_1 = \frac{\Delta Q_x p c}{\Delta Q_L} \quad (84)$$

This complex number measures the coherent tune shift in total incoherent tune spread units. The integral can then be written:

$$J_{\perp} = \int_0^1 \frac{x \, dx}{x - x_1} \quad (85)$$

We also note

$$\Lambda_{\perp pc} = \frac{-\left(\frac{q}{A}\right)I_c}{2\pi Q_x \omega_0 \left(\frac{m_0 c^2}{e}\right) \gamma_0 \Delta Q_L} \quad (86)$$

and are left with

$$\frac{1}{J_{\perp}} = \Lambda_{\perp} pc j Z_{\perp}(p) \quad (87)$$

As was already done in the previous examples, the solutions can be worked out by drawing the curves corresponding to a given value of  $\text{Im}(x_1)$  in the stability diagram.

$$\begin{aligned} \Lambda_{\perp} pc \text{Re}(Z_{\perp}(p)) &= \text{Im}\left(\frac{1}{J_{\perp}}\right) \\ \Lambda_{\perp} pc \text{Im}(Z_{\perp}(p)) &= -\text{Re}\left(\frac{1}{J_{\perp}}\right) \end{aligned} \quad (88)$$

The results are shown in Fig. 17.

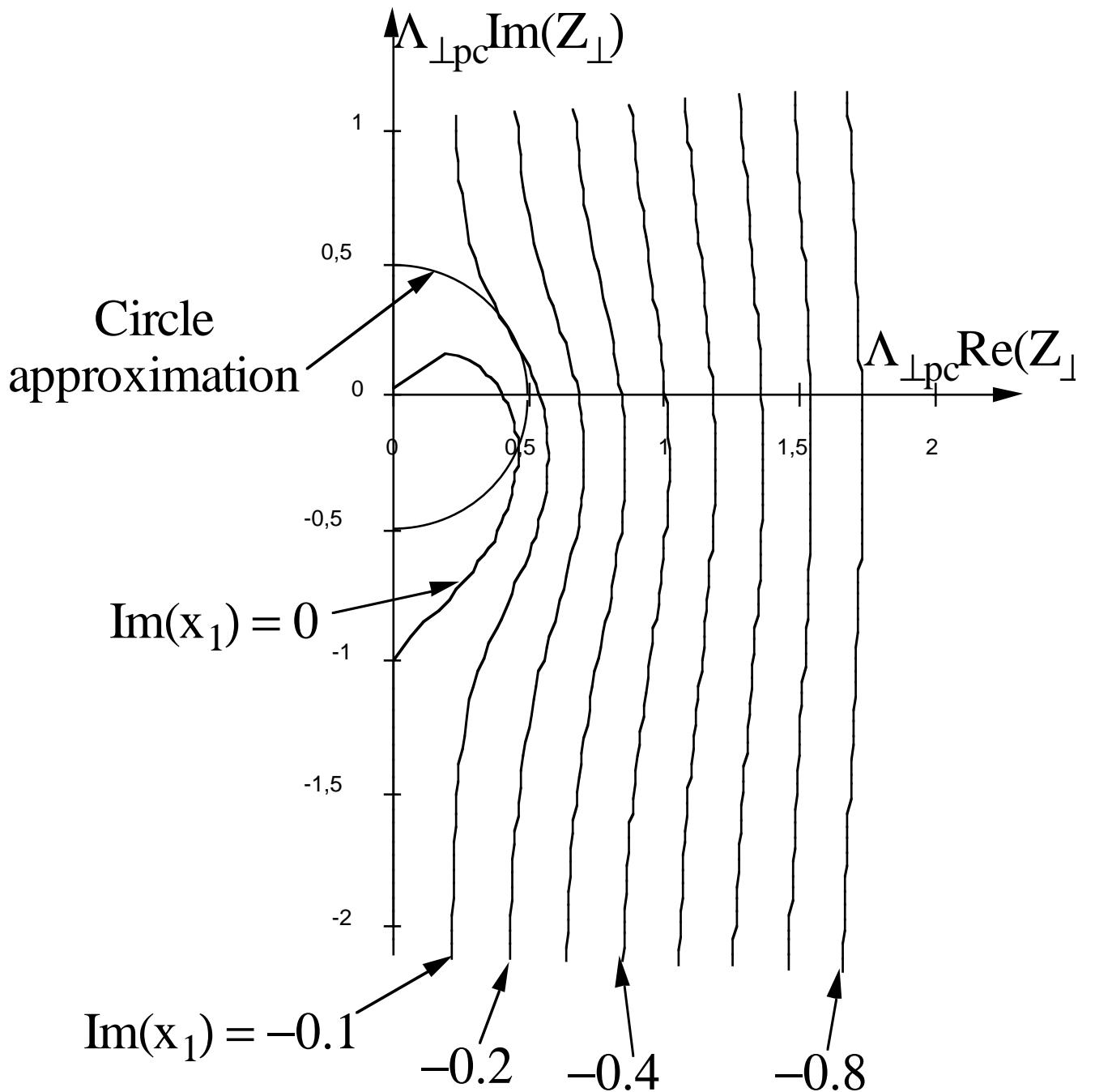


Fig . 17

As already mentioned, the details of the stability contour depend on the tails of the  $\frac{\partial f_0}{\partial \hat{x}}$

distribution. The transverse distribution enters the integral via its derivative  $\frac{\partial \hat{x}}{\partial \hat{x}}$  which for the specific suggested example presents a discontinuity at the edges. This explains the heart shape of the stability contour which would be much less pronounced, and even not exist, for a smoother distribution. In view of this, we suggest using

$$|\Lambda_{\perp pc} Z_{\perp}(p)| < 0.5 \quad (89)$$

as an approximation of the stability criterion. The interpretation is very similar to that

given in the previous paragraph for a momentum spread. When the incoherent tune spread due to amplitude spread is large enough,

$$\Delta Q_L \omega_0 = \Delta \omega_p L > \frac{\left(\frac{q}{A}\right) I_c |Z_{\perp}(p)|}{\pi Q_{x0} \left(\frac{m_0 c^2}{e}\right) \gamma_0} \quad (90)$$

coherent motion is Landau damped.

If one compares the above result with that previously obtained for a momentum spread,

$$\Delta \omega_p \text{FWHH} > \frac{\left(\frac{q}{A}\right) I_c |Z_{\perp}(p)|}{4 Q_{x0} \left(\frac{m_0 c^2}{e}\right) \gamma_0} \quad (91)$$

$$\text{with } \Delta \omega_p \text{FWHH} = \left( (p - Q_{x0}) \omega_0 + \omega_{\xi} \right) \left| \eta \left( \frac{\delta p}{p} \right) \text{FWHH} \right| \quad (92)$$

apart from the distinction between the definitions used for FWHH and total (L), the main difference is that the width of the incoherent band due to amplitude spread is now independent of p. Another way to summarize the results is to express the necessary incoherent spread in terms of the coherent betatron frequency shift one would obtain with a monochromatic beam.

$$\Delta \omega_p L > 4 \omega_0 |Q_{xpc} - Q_{x0}|_{\text{monochromatic beam}} \quad (93)$$

## BIBLIOGRAPHY

- C.E. Nielsen et al., Proc. of Int. Conf. on High Energy Accel. and Instr., CERN (1959) 239
- V.K. Neil, R.J. Briggs, Plasma Physics, 9 (1967) 631
- V.K. Neil, A.M. Sessler, Rev. Sci. Instr. 36 (1965) 429
- B.W. Montague, Single Particle Dynamics, 3rd lecture, Int. School of Particle Accel., Erice (1976)
- B.W. Montague, Single Particle Dynamics, 5th and 6th lectures, Int. School of Particle Accel., Erice (1976)
- H.G. Hereward, Landau Damping, Int. School of Particle Accel., Erice (1976)
- K. Hübner, V.G. Vaccaro, CERN report ISR-TH/70-44 (1970)
- B. Zotter, CERN report ISR-GS/76-11 (1976)
- A.G. Ruggiero, V.G. Vaccaro, CERN report ISR-TH/68-33 (1968)
- E. Keil, W. Schnell, CERN report ISR-TH-RF/69-48 (1969)
- F. Sacherer, Proc. 1973 Part. Accel. Conf., San Francisco, IEEE Trans. Nucl. Sci. Vol. NS-20, n°3, 825
- F. Sacherer, CERN report SI/BR/72-5 (1972)
- A. Hofmann, Single Beam collective phenomena-longitudinal, Int. School of Part. Accel., Erice (1976)
- I. Hofmann, Non linear aspects of Landau damping in computer simulation of the microwave instability, Computing in Accelerator Design and Operation, Berlin, Sept. 20-23 1983
- J.L. Laclare, Bunched-beam instabilities, 11th Int. conference on High Energy Accelerators, Geneva (July 1980) 526
- J.L. Laclare, Instabilities in storage rings, Proc. of the Symposium on Accelerator Aspects of H I Fusion, Darmstadt FRG (1982), GSI Report 82-8.

# SINGLE-BEAM LIFETIME

*A. Wrulich*

Sincrotrone Trieste, Italy

## Abstract

An overview of the most relevant single beam effects, which cause a reduction in beam lifetime in storage rings, is given. The considerations, however, are limited to perturbations which have a random time character. After defining the aperture limitations which are responsible for the reduction in beam intensity, the lifetime due to quantum fluctuation and radiation damping for electrons (or positrons) is derived, followed by a treatment of the intensity reduction due to statistical fluctuations without damping. Furthermore the effects of beam gas scattering, the effects of Coulomb scattering inside a bunch (i.e. Touschek scattering and multiple intrabeam scattering) and the effects of resonance crossing on the lifetime are summarized.

## 1. INTRODUCTION

To give a general definition, the lifetime of a single beam is the time a beam can be kept, usually in a storage ring. Consequently all effects which reduce the beam intensity should be taken in account. The considerations given here, however, will be limited to the reduction of beam intensity caused by small perturbations which have a random time dependence. In this sense, lifetime is usually defined as the time the beam intensity decays to a certain fraction of its initial value (half or  $1/e$ ).

Such effects are for instance:

- i) Effect of noise on betatron oscillation and synchrotron oscillation;
- ii) Scattering by residual gas (elastic, inelastic);
- iii) Quantum fluctuation of radiation;
- iv) Scattering between particles inside a bunch (IBS, Touschek scattering);
- v) Multiple traversal of resonances;

### 1.1 Aperture Limitations

The lifetime given by all these effects is related to the final aperture available for the particle motion. Without this limitation the average intensity, i.e. average stored current would remain constant (usually used to define the lifetime). Only the current densities would be changed and the bunch would become longer and wider and might not be useful anymore for experiments (the luminosity for colliders would be decreased and the brilliance for radiation sources would be reduced).

Aperture limitation, does not necessarily mean a physical limitation, but also the limitation in the transverse plane due to the dynamic aperture, or the limitation in the longitudinal plane due to the RF-bucket size or the dynamic aperture for off-momentum particles (i.e. particles with large synchrotron oscillation amplitudes).

#### 1.1.1 Transverse Limitation

In the linear treatment of betatron oscillation the particle is lost, if its amplitude exceeds the aperture of the vacuum chamber.

Normally, non-linear magnetic fields present in the accelerator (as for instance sextupoles which are introduced to compensate chromatic effects) cause a limitation of the

maximum betatron amplitude, described by the dynamic aperture. Also in this case the particle is lost at the physical aperture, but non-linear effects blow up the betatron motion and limit the "stable" initial amplitudes to values far below the physical aperture.

Figure 1 shows the relation between initial amplitude and maximum final amplitude due to the non-linear effects. The initial amplitude for which the final one approaches infinity is usually called the dynamic aperture.

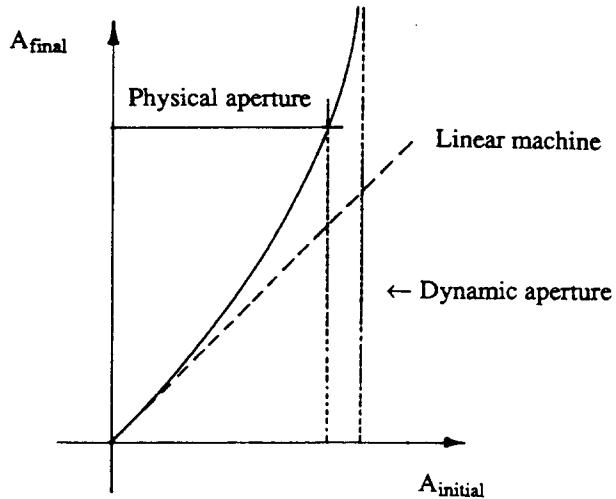


Fig. 1 Dynamic aperture reductions due to non-linearities. The increased betatron amplitude by non-linearities is drawn as a function of the initial amplitude.

### 1.1.2 Longitudinal Limitation

In the longitudinal plane the particle is lost either at the RF-acceptance limit or the momentum acceptance of the dynamic aperture.

#### RF acceptance

This is the limit of self focusing for the longitudinal motion, which is described by the separatrix in longitudinal phase space (see figure 2).

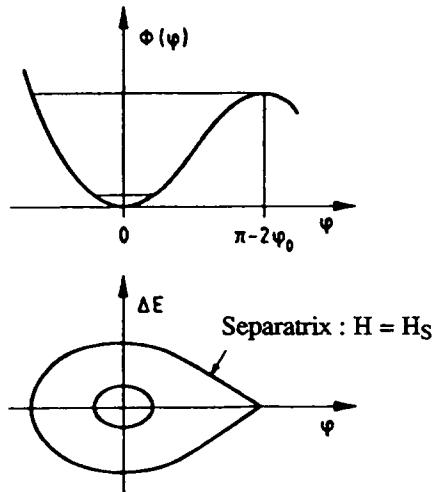


Fig. 2 Aperture limitation in the longitudinal plane due to the self-focusing of the RF.  
Potential and phase space separatrix are shown.

with:

$$H_s = \frac{eU}{\pi h \alpha_c E} \left[ \cos\phi_0 + \left( \phi_0 - \frac{\pi}{2} \right) \sin\phi \right] \quad (1)$$

and the meaning:

- $U$  RF voltage,
- $\phi_0$  synchronous phase,
- $h$  harmonic number,
- $E$  total particle energy,
- $\alpha_c$  momentum compaction factor.

#### *Momentum acceptance of dynamic aperture*

For particles with momentum deviation, the dynamic aperture can be strongly reduced as compared to the zero momentum case. Figure 3 shows the dynamic aperture reduction in ELETTRA with increasing momentum deviation, i.e. synchrotron oscillation amplitude.

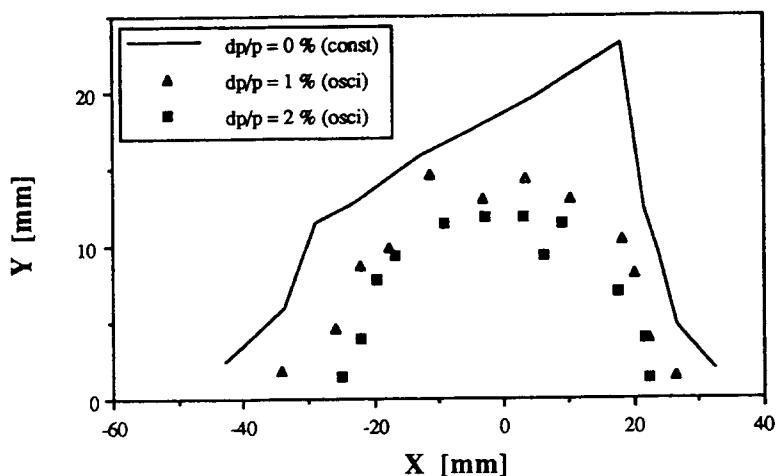


Fig. 3 Reduction in off-momentum aperture for ELETTRA.

## 1.2 Particle Distribution in a Storage Ring in the Presence of Small Statistically Independent Perturbations

If we want to evaluate the fraction of particles lost by a limitation in aperture, we have to know the distribution function, i.e. what part of the distribution is cut off by the aperture limitation.

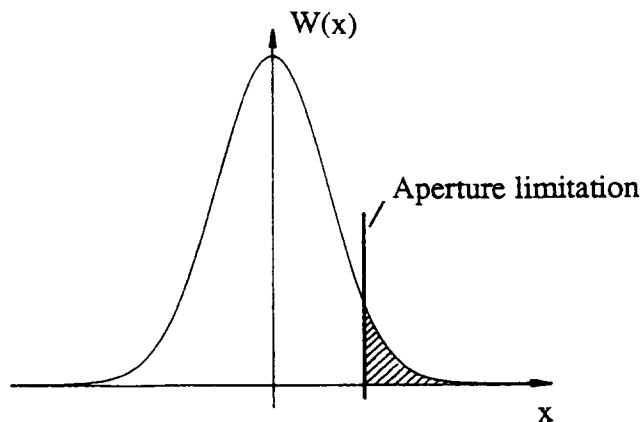


Fig. 4 Sketch of particle loss of a distribution function due to an aperture limitation.

The particles from the shadowed area in figure 4 are lost by the aperture limitation. For an instantaneous limitation one has an initial intensity drop, but then the intensity remains constant, if now diffusion process exists to refill the shadowed area.

The effect of statistically independent perturbations can be described by the Fokker-Planck equation [1,2]:

$$\frac{\partial w(x,t)}{\partial t} = -\frac{\partial}{\partial x} [\langle x \rangle w(x,t)] + \frac{\partial^2}{\partial x^2} \left[ \left( \frac{\langle x^2 \rangle}{2} \right) w(x,t) \right] \quad (2)$$

where  $w$  is the distribution function,  $\langle x \rangle$  is the mean increase per unit time and  $\langle x^2 \rangle$  is the mean square of the increase per unit time. Physically the Fokker-Planck equation is a continuity equation:

$$\frac{\partial w}{\partial t} = \text{div } I = -\frac{\partial}{\partial x} \left[ \langle x \rangle w - \frac{\partial}{\partial x} \left( \frac{\langle x^2 \rangle}{2} \right) w \right] \quad (3)$$

with  $I$  the current along the  $x$ -axis. The first term represents a systematic component corresponding to the mean displacement of the velocity (i.e. damping or excitation) and the second term describes a diffusion component proportional to the gradient of the concentration.

## 2. LIFETIME DUE TO QUANTUM FLUCTUATION AND RADIATION DAMPING

Lifetime effects due to quantum fluctuations and radiation damping are essentially only important for electrons [3], since the radiation effects are small for higher particle masses and therefore usually insignificant for protons. Although protons in the SSC, at 20 TeV reveal a decent radiation damping effect, very welcome there to combat beam growth due to intrabeam scattering.

### 2.1 Transverse Plane

The Fokker-Planck equation for the distribution of the emittance becomes [4]:

$$\frac{\partial w(\varepsilon,t)}{\partial t} = -\frac{\partial}{\partial \varepsilon} \left\{ \langle \delta \varepsilon \rangle w(\varepsilon,t) + \frac{\partial}{\partial \varepsilon} \left[ \left( \frac{\langle \delta^2 \varepsilon \rangle}{2} \right) w(\varepsilon,t) \right] \right\} \quad (4)$$

where  $\langle \delta \varepsilon \rangle$  is the mean change of the emittance per unit time and  $\langle \delta^2 \varepsilon \rangle$  is the mean square of the change per unit time. The equation can be solved, after these quantities have been derived. There are two effects contributing to these parameters. First the statistical nature of the quantum emission which increases the beam size (excitation), and then the restoring of the energy in the accelerating cavities, which acts to decrease the beam dimensions (damping).

If a radiation quantum is emitted, the energy of the particle is reduced, which changes the corresponding closed orbit (particles with different energies have different closed orbits, as described by the dispersion). If a quantum with energy  $\delta E$  is emitted, the betatron oscillation coordinates are changed according to:

$$x \rightarrow x + D \frac{\delta E}{E}$$

$$x' \rightarrow x' + D' \frac{\delta E}{E}$$

substituting these changes in the expression for the emittance:

$$\epsilon\beta = \left( x + D \frac{\delta E}{E} \right)^2 + \left[ (x + D \frac{\delta E}{E}) \alpha + (x' + D' \frac{\delta E}{E}) \beta \right]^2 \quad (5)$$

we find after some manipulations:

$$\begin{aligned} \langle \delta \epsilon \rangle &= -2 a_\beta \epsilon + M \\ \langle \delta \epsilon^2 \rangle &= 2 \epsilon M \end{aligned}$$

with "a<sub>β</sub>" the damping decrement, which is made up by the excitation term due to quantum emission and the damping term created by the recompensation of the energy in the accelerating cavities.  $M = H \frac{\delta E^2}{E}$ , with the Courant Snyder invariant  $H = D^2\gamma + 2DD'\alpha + D'^2\beta$ .

After substitution of these quantities the Fokker-Planck equation becomes:

$$\frac{\partial w(\epsilon, t)}{\partial t} = \frac{\partial}{\partial \epsilon} \left[ 2 \epsilon a_\beta w(\epsilon, t) + \frac{\epsilon M \partial w(\epsilon, t)}{\partial \epsilon} \right] \quad (6)$$

for the stationary distribution the time derivative vanishes and we get the solution:

$$w_\epsilon(\epsilon) = \frac{2a_\beta}{M} e^{-\frac{2a_\beta}{M}\epsilon} \quad (7)$$

Replacing the emittance  $\epsilon$  by the transverse coordinate  $x$ , we find the well-known gaussian distribution for the transverse beam dimensions due to betatron oscillations:

$$w_x(x) = \frac{1}{\sqrt{2\pi\sigma_x^2}} e^{-\frac{x^2}{2\sigma_x^2}} \quad (8)$$

Since the distribution extends to infinity, the vacuum chamber walls will cause a loss of particles. Assuming a maximum value of the emittance  $\epsilon_m$  due to aperture limitations, the solution of the Fokker-Planck equation must now vanish at this boundary:

$$w_\epsilon(\epsilon_m) = 0$$

In the case of a limitation of the emittance by  $\epsilon_m$ , the time dependence of the distribution function can be separated as:

$$w(\epsilon, t) = e^{-\frac{t}{\tau}} \bar{w}(\epsilon) \quad (9)$$

$$-\frac{\partial \bar{w}(\epsilon)}{\partial \epsilon} = \frac{\partial}{\partial \epsilon} \left[ 2 \epsilon a_\beta \bar{w}(\epsilon) + \frac{\epsilon M \partial \bar{w}(\epsilon)}{\partial \epsilon} \right] \quad (10)$$

After performing the integration  $\int \dots d\epsilon$ , we find:

$$-\frac{1}{\tau} = \epsilon_m M \left. \frac{\partial \bar{w}(\epsilon)}{\partial \epsilon} \right|_{\epsilon_m} \quad (11)$$

if we substitute here as an approximation the solution without limitation, we get for the lifetime:

$$\tau = \frac{1}{2 a_\beta r_\beta} e^{r_\beta} \quad \text{with} \quad r_\beta = \frac{2 a_\beta}{M} \quad \epsilon_m = \frac{1}{2} \left( \frac{x_\beta}{\sigma_\beta} \right)^2$$

The equation can also be solved more accurately by using a series expansion and solving [5]:

$$\bar{w}'' \epsilon M + \bar{w}' (2a\epsilon + M) + \left( 2a + \frac{1}{\tau} \right) \bar{w} = 0 \quad (12)$$

## 2.2 Longitudinal Plane

The same equation derived for the longitudinal plane (assuming that the limitation in momentum is given by the RF-bucket size) leads to:

$$\tau = \frac{1}{2 a_s r_s} e^{r_s} \quad \text{with} \quad r_s = \frac{1}{2} \left( \frac{\Delta E}{\sigma_E} \right)^2$$

where  $\Delta E$  is the energy acceptance of the bucket provided by the RF-system and  $\sigma_E$  the standard deviation of the energy distribution.

## 3. LIFETIME FOR STATISTICAL FLUCTUATIONS WITHOUT DAMPING

If the statistical perturbation is independent of the variable the mean value does not contain a damping term and the Fokker-Planck equation (6) reduces to a pure diffusion equation. Such random perturbations are for instance caused by fluctuations of the magnetic field or scattering on residual gas particles:

$$\frac{\partial w}{\partial t} = M \frac{\partial}{\partial \epsilon} \left( \epsilon \frac{\partial w}{\partial \epsilon} \right) \quad (13)$$

we assume that the motion is limited by a maximum value of the emittance  $\epsilon_m$ . A change in variables of the form:

$$\tau = \left( \frac{M}{\epsilon_m} \right) t$$

and

$$\theta = \frac{\epsilon}{\epsilon_m}$$

simplifies the equation to:

$$\frac{\partial w}{\partial \tau} = \frac{\partial}{\partial \theta} \left( \frac{\partial w}{\partial \theta} \right) \quad (14)$$

With the initial condition and boundary condition respectively:

$$\begin{aligned} w(\theta, 0) &= w_0(\theta) \\ w(1, \tau) &= 0 \end{aligned}$$

the differential equation has solutions of the form [1,6]:

$$w(\theta, \tau) = \sum_n C_n J_0(\lambda_n \sqrt{\theta}) e^{-\frac{\lambda_n^2 \tau}{4}} \quad (15)$$

with the coefficients:

$$C_n = \frac{1}{J_1(\lambda_n)^2} \int_0^1 w_0(\theta) J_0(\lambda_n \sqrt{\theta}) d\theta \quad (16)$$

where  $\lambda_n$  is the n'th zero of the Bessel function  $J_0(\theta)$ .

Assuming as previously an initial emittance distribution which corresponds to a Gaussian distribution in betatron coordinates, we have:

$$w_0(\varepsilon) = \frac{1}{2\varepsilon_0} e^{-\frac{\varepsilon}{2\varepsilon_0}} \quad (17)$$

or in the new coordinates:

$$w_0(\theta) d\theta = \frac{\varepsilon_m}{2\varepsilon_0} e^{-\frac{\varepsilon_m}{2\varepsilon_0} \theta} d\theta \quad (18)$$

where  $\varepsilon_0$  is the emittance corresponding to one sigma of the transverse phase space dimensions. We then find for the coefficients:

$$C_n = \frac{\varepsilon_m}{2\varepsilon_0} \frac{1}{J_1(\lambda_n)^2} \int_0^1 e^{-\left(\frac{\varepsilon_m}{2\varepsilon_0}\right)\theta} J_0(\lambda_n \sqrt{\theta}) d\theta \quad (19)$$

Generally the integration must be done numerically. Only if  $\varepsilon_m$  is large, can the integration be performed and we get for the coefficients:

$$C_n = \frac{1}{J_1(\lambda_n)^2} e^{-\frac{\lambda_n^2}{2} \frac{\varepsilon_0}{\varepsilon_m}} \quad (20)$$

For the development of the beam intensity in time, we get (for large  $\varepsilon_m$ ):

$$\begin{aligned} N(\tau) &= \int_0^1 w(\theta, \tau) d\theta \\ &= 2 \sum_n \frac{1}{\lambda_n J_1(\lambda_n)} e^{-\lambda_n^2 \left[ \frac{\tau}{4} + \frac{\varepsilon_0}{\varepsilon_m} \right]} \end{aligned} \quad (21)$$

The lifetime is defined as:

$$T = -\frac{N(\tau)}{dN/d\tau} \quad (22)$$

which reaches an asymptotic value given by:

$$T_\infty = \frac{\lambda_1^2}{4} \frac{\varepsilon_m}{M} \quad (23)$$

Figure 5 shows the variation of the lifetime for two different parameters of  $\left(\frac{\epsilon_0}{\epsilon_m}\right)$ .

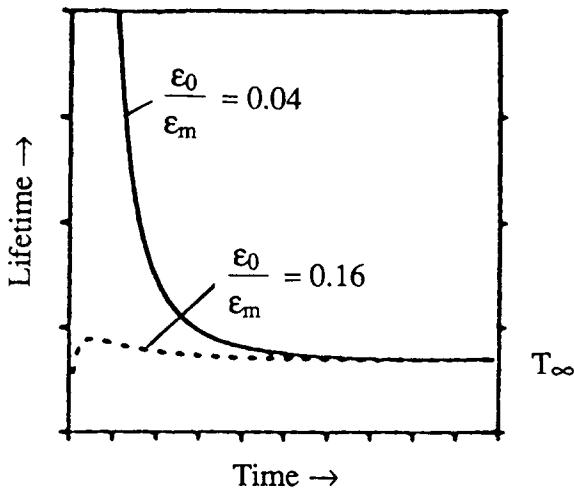


Fig. 5 Change in lifetime for random perturbations without damping, for two different aperture limitations. For both initial conditions, the lifetime is approaching the asymptotic lifetime [6].

For small values, i.e. larger aperture limits, the lifetime is long at the beginning and decreases slowly to the asymptotic value. If the aperture is small, the initial lifetime is very short and then approaches rapidly the asymptotic value.

#### 4 LIFETIME DUE TO BEAM-GAS SCATTERING

Particles are lost by scattering at the residual gas molecules. This effect can easily be controlled by providing sufficient pumping to reach a low residual gas pressure. There are two principally different effects:

*Elastic Scattering:* where the stored particle is transversally deflected and increases its betatron oscillation amplitude. If the change is large enough, the particle is lost at either the physical aperture or the dynamic aperture.

*Inelastic Scattering:* besides the deflection, where the particle does not loose energy, there is also the possibility that a light quantum is emitted during the collision and the energy of the particle is changed or the particle transfers energy to the atom of the rest gas.  
In both cases the particle looses energy and consequently gets lost at the RF acceptance limit or the off-momentum dynamic aperture limit.

##### 4.1 Elastic Scattering [7]

An incoming homogeneous particle current with constant velocity is scattered at a point-like Coulomb field given by the charge of the nucleus of the residual gas atom.

###### 4.1.1 Rutherford Scattering [8]

We are interested in the cross section of the scattering effect which is easily derived from the graph in figure 6:

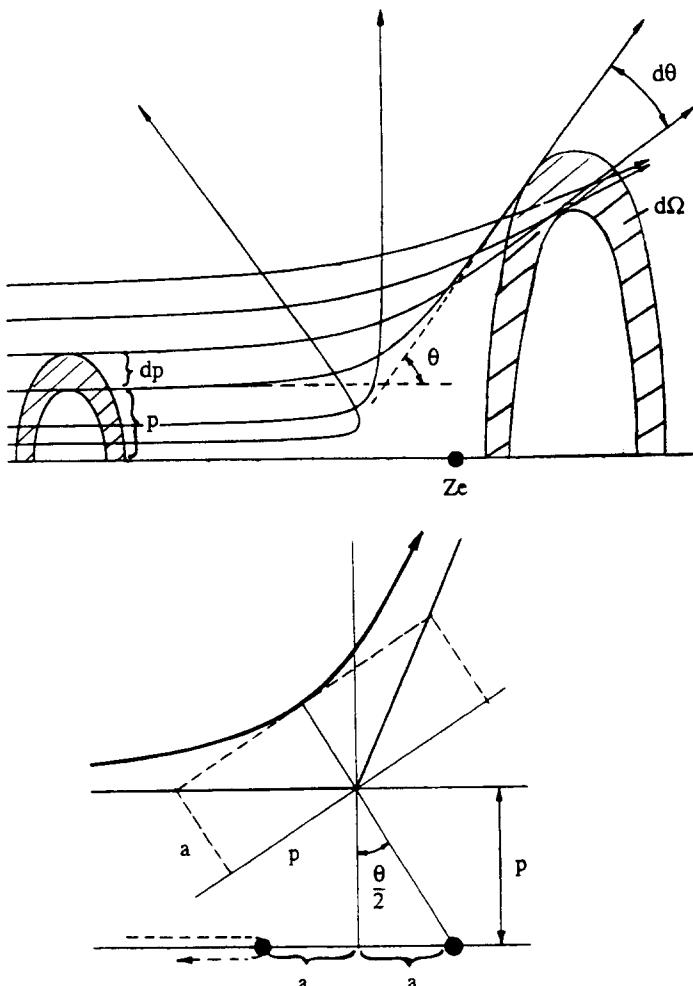


Fig. 6 Scattering of an electron on a nucleus (Rutherford scattering) and geometry of the scattering process.

$$\text{Geometry: } \tan \frac{\theta}{2} = \frac{a}{p} \quad (24)$$

All scattering curves are hyperbolas with the impact parameter "p" for half of the main axis and "a" the half-size of the secondary axis.

The quantity "a" can be easily defined by looking at a particle which moves on axis in the direction of the Coulomb center. It is reflected at a position in which the kinetic energy equals the potential energy:

$$mc^2 = \frac{Ze^2}{2a} \rightarrow a = \frac{Ze^2}{2m_0 c^2 \gamma} \quad (25)$$

The incoming ring  $d\sigma$  of particles is scattered in the solid angle  $d\Omega$ .

$$\begin{aligned} d\sigma &= 2\pi p dp \\ d\Omega &= 2\pi \sin \theta d\theta \end{aligned} \Rightarrow d\sigma = 2\pi \left( \frac{a}{\tan \theta/2} \right) d \left( \frac{a}{\tan \theta/2} \right) = \frac{a^2}{2} \frac{d\Omega}{\sin^4 \theta/2} \quad (26)$$

In the small angle approximation, which is valid for all practical cases we are interested in, we get for the differential cross section:

$$\frac{d\sigma}{d\Omega} = \left( \frac{Ze^2}{m_0 c^2 \gamma} \right)^2 \frac{2}{\theta^4} \quad (27)$$

The cross section has to be corrected for two extreme cases:

For small angles, i.e. large deviations,  $p$  from the central axis, the screening effect of the electrons on the atom has to be taken into account, which causes a fall off of the potential more rapidly than  $\frac{1}{r}$ .

At large angles the cross section is changed due to the finite size of the nucleus. Figure 7 shows these corrections on the cross section given by eq. (27).

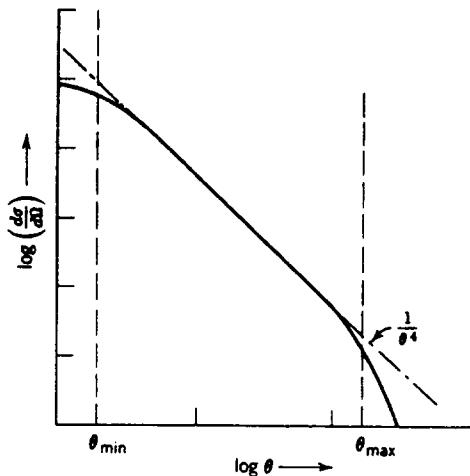


Fig. 7 Change of the differential cross section for the effects of electron screening at small angles and finite nuclear size at large angles [8].

The cross section flattens off at small angles to a finite value at  $\theta = 0$ . The modified cross section has the form:

$$\frac{d\sigma}{d\Omega} = \left( \frac{Ze^2 2}{m_0 c^2 \gamma} \right) \frac{1}{(\theta^2 + \theta_1^2)^2} \quad (28)$$

where  $\theta_1$ , the minimum angle due to electron shielding.

Integration between  $\theta_0$ , the minimum angle for which a loss occurs and  $\theta_{\max}$  (which we don't need to specify, because the dominant losses happen at small angles) leads to:

$$\sigma = \frac{2\pi Z^2 r_0^2}{\gamma^2} \left\{ \frac{1}{\theta_0^2 + \theta_1^2} - \frac{1}{\theta_{\max}^2 + \theta_1^2} \right\} \quad (29)$$

$$\theta_1 = \frac{Z^{1/3}}{192} \cdot \frac{1}{\gamma} \quad (30)$$

which is approximately:

$$\sigma = \frac{2\pi Z^2 r_0^2}{\gamma^2} \cdot \frac{1}{\theta_0^2} \quad (31)$$

The maximum betatron amplitude generated by a deflection  $\theta_0$  is given by:

$$A_i = \theta_0 \sqrt{\beta_i} \sqrt{\hat{\beta}} \quad (32)$$

The particle is lost if the amplitude reaches the half aperture  $H$  or the dynamic limit, i.e.:

$$A_i \geq H \quad (33)$$

averaged over all positions of the ring, we get for the cross section:

$$\sigma = \frac{2\pi Z^2 r_0^2}{\gamma^2} \cdot \frac{\bar{\beta} \hat{\beta}}{H^2} \quad (34)$$

Accordingly the cross section becomes smaller for an optics with small beta functions and for increasing machine energies.

## 4.2 Inelastic Scattering [7,9]

We have to deal here with two effects:

- i) Bremsstrahlung scattering, where the particle emits a photon (and the nucleus is left unexcited).
- ii) Inelastic scattering from an electron of the atom in which the momentum transfer excites the atom.

(There is also an elastic scattering at the electrons, but its effect is negligible).

### 4.2.1 Bremsstrahlung scattering at the nucleus [10]

$$\left( \frac{d\sigma}{d\epsilon} \right)_A = \alpha \frac{Z^2 r_0^2}{\epsilon} \left\{ \left[ \frac{4}{3} \left( 1 - \frac{\epsilon}{E} \right) + \frac{\epsilon^2}{E^2} \right] \left[ \Phi_1(\epsilon_1) - \frac{4}{3} \ln Z \right] \right\} \\ + \left[ \frac{2}{3} \left( 1 - \frac{\epsilon}{E} \right) \right] [\Phi_2(\epsilon_1) - \Phi_1(\epsilon_1)] \quad (35)$$

with:  $\alpha = \frac{e^2}{hc} = \frac{1}{137}$

$\epsilon$  = photon energy,

$E$  = particle energy,

$\Phi_1, \Phi_2$  = screening functions.

For high energy electrons the screening of the nucleus by the outer electrons must be taken into account. The argument of the screening functions  $\Phi_1, \Phi_2$ , is given by:

$$\epsilon_1 = 100 \frac{\epsilon}{Z^{1/3}} \cdot \frac{m_0 c^2}{E(E-\epsilon)} \quad (36)$$

and can be approximated by  $\epsilon \approx 0$  for high energies, which correspond to the case of complete screening. In this case we get for the screening functions:

$$\begin{aligned}\varphi_1(0) &= 20.836 \\ \varphi_2(0) - \varphi_1(0) &= \frac{2}{3}\end{aligned}$$

and for the differential cross section:

$$\left(\frac{d\sigma}{d\epsilon}\right)_A = \alpha \frac{4Z^2 r_0^2}{\epsilon} \left\{ \left[ \frac{4}{3} \left(1 - \frac{\epsilon}{E}\right) + \frac{\epsilon^2}{E^2} \right] \left[ \frac{\varphi_1(0)}{4} - \frac{1}{3} \ln Z \right] \right. \\ \left. + \left[ \frac{1}{9} \left(1 - \frac{\epsilon}{E}\right) \right] \right\} \quad (37)$$

#### 4.2.2 Inelastic scattering off the electrons of the residual gas atom [11]

$$\left(\frac{d\sigma}{d\epsilon}\right)_B = \alpha \frac{Z r_0^2}{\epsilon} \left\{ \left[ \frac{4}{3} \left(1 - \frac{\epsilon}{E}\right) + \frac{\epsilon^2}{E^2} \right] \left[ \psi_1(\epsilon_2) - \frac{8}{3} \ln Z \right] \right. \\ \left. + \left[ \frac{2}{3} \left(1 - \frac{\epsilon}{E}\right) \right] [\psi_2(\epsilon_2) - \psi_1(\epsilon_1)] \right\} \quad (38)$$

The functions  $\psi_1, \psi_2$  have the same meaning as  $\varphi_1, \varphi_2$  before.

For:

$$\epsilon_2 = 100 \frac{\epsilon}{Z^{2/3}} \frac{mc^2}{E(E-\epsilon)} \approx 0 \quad (39)$$

we find:

$$\psi_1(0) = 28.34, \psi_2(0) - \psi_1(0) = \frac{2}{3} \quad (40)$$

and:

$$\left(\frac{d\sigma}{d\epsilon}\right)_B = \alpha \frac{4Z r_0^2}{\epsilon} \left\{ \left[ \frac{4}{3} \left(1 - \frac{\epsilon}{E}\right) + \frac{\epsilon^2}{E^2} \right] \left[ \frac{\psi_1(0)}{4} - \frac{2}{3} \ln Z \right] + \left[ \frac{1}{9} \left(1 - \frac{\epsilon}{E}\right) \right] \right\} \quad (41)$$

The total differential cross section is given by the sum of both contributions. We can now derive the cross section of inelastic scattering in terms of  $\epsilon_m$  which is the lower boundary for which the particles start to be lost:

$$\sigma = \int_{\epsilon_m}^E \left(\frac{d\sigma}{d\epsilon}\right) d\epsilon = 4 \alpha r_0^2 \left\{ F \int \left[ \frac{4}{3\epsilon} - \frac{4}{3E} + \frac{\epsilon}{E^2} \right] d\epsilon \right. \\ \left. + \frac{1}{9} Z (Z+1) \int \left[ \frac{1}{\epsilon} - \frac{1}{E} \right] d\epsilon \right\} \quad (42)$$

if  $\epsilon_m \ll E$  we find:

$$\sigma = 4\alpha r_0^2 \left\{ F \frac{4}{3} \left[ \ln \frac{E}{\epsilon_m} - \frac{5}{8} \right] + \frac{1}{9} Z (Z+1) \left[ \ln \frac{E}{\epsilon_m} - 1 \right] \right\} \quad (43)$$

with:

$$\begin{aligned}
 F &= Z^2 \left[ \frac{\varphi_1(0)}{4} - \frac{1}{3} \ln Z \right] + Z \left[ \frac{\psi_1(0)}{4} - \frac{2}{3} \ln Z \right] \\
 &= Z^2 \ln \frac{e \frac{\varphi_1(0)}{4}}{Z^{1/3}} + Z \ln \frac{e \frac{\psi_1(0)}{4}}{Z^{2/3}} \\
 &= Z^2 \ln \frac{183}{Z^{1/3}} + Z \ln \frac{1194}{Z^{2/3}}
 \end{aligned} \tag{44}$$

For most of the practical cases the contribution from the electrons is negligible.

### 4.3 Beam-Gas Scattering Lifetime

The electrons lost after passing through a volume with residual gas atoms is given by:

$$dn = -n_0 \sigma N dx \tag{45}$$

which leads to the loss rate:

$$\frac{1}{\tau} = -\frac{1}{n_0} \frac{dn}{dt} = c \sigma N$$

with  $n_0$  the number of electrons and  $N$  the gas atoms per unit volume.

If desorption effects from the vacuum chamber walls are still relevant, the number of residual gas particles depends also on the number of electrons,  $n_0$ .

With:

$$N = N_0 + G \cdot n_0$$

we find:

$$\frac{dn}{dt} = -\sigma c [n_0 N_0 + n_0^2 G]$$

Defining the lifetime as the time it takes for the initial particle intensity to be reduced by  $\frac{1}{e}$ , we get:

$$\frac{1}{\tau} = (\sigma c N_0) \ln \frac{N_0 + G n_0}{e N_0 + G n_0}$$

which reduces to the previous expansion if  $G = 0$ , i.e.:

$$\frac{1}{\tau} = (\sigma c) N_0$$

For  $N_0 = 0$ , the vacuum is dominated by the desorption effects, we get:

$$\frac{1}{\tau} = \frac{(\sigma c) G n_0}{e-1}$$

Since the residual gas is made up of different molecules each of which may contain more than one atom, we must replace  $N$  by the sum:

$$N = \sum_{i,j} k_{i,j} N_i \tag{46}$$

$k_{ij}$  = number of atoms (j) in the molecule (n)  
 $N_i$  = number of molecules of type (i)

With  $N_i = \frac{p_i}{kT}$  we get for the lifetime (neglecting desorption effects):

$$\frac{1}{\tau} = - \frac{1}{n_0} \frac{dn}{dt} = \frac{c\sigma}{kT} \cdot \sum_{i,j} k_{i,j} p_i \quad (47)$$

where  $p_i$  is the partial pressure for gas species  $i$ .

#### 4.4 Lifetime Reduction Due to Ion Trapping [12-14]

So far we have only considered the normal pressure given by the residual gas in the vacuum chamber. But there are also effects which can locally enhance the gas pressure inside the volume of the circulating beam. This mechanism is called ion trapping and is only relevant to electron machines.

An ion created by a collision of the electron with the rest gas, gets affected by the electromagnetic force of the circulating bunches, which impose a transverse focusing on the ion, given by:

$$\begin{aligned} \Delta x' &= a_x \cdot x \\ \Delta y' &= a_y \cdot y \end{aligned} \quad (48)$$

(For larger ion displacements  $x,y$  the expressions are not valid anymore and the force becomes highly non-linear.) The focusing parameter is given by:

$$a_{x,y} = \frac{2 r_p c N_B}{A \sigma_{x,y} (\sigma_x + \sigma_y)} \quad (49)$$

with  $A$  the atomic mass number,  $N_B$  the number of electrons per bunch and  $\sigma_{x,y}$  the rms beam sizes.

The principles for linear stability can now be applied for the small displacement approximation. For a beam with equally spaced bunches and identical particle population, one period of the force acting on the ions is given by the kick described above, followed by a drift equal to the bunch separation, i.e.:

$$M = \begin{bmatrix} 1 & t_B \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -a & 1 \end{bmatrix} \quad (50)$$

with  $t_B$  the time separation between two bunches and "a" the focusing parameter. The motion is stable, i.e. ions are trapped, if:

$$-2 < T_r(M) = 2 - a t_B < 2 \quad (51)$$

This limit is reached for a critical ion mass:

$$(A_c)_{x,y} = \frac{r_p N_B s_B}{2 \sigma_{x,y} (\sigma_x + \sigma_y)} \quad (52)$$

with  $s_B$  the bunch distance. All ion masses above the critical mass are trapped.

The ion accumulation continues until defocusing forces due to the ion space charge in the drift region are strong enough to establish an equilibrium. The drift region has to be replaced now by the defocusing transfer matrix which is given by:

$$M_{ion} = \begin{bmatrix} \cosh(gt) & \frac{1}{g} \sinh(gt) \\ g \sinh(gt) & \cosh(gt) \end{bmatrix} \quad (53)$$

with:

$$g^2 = \frac{e}{Am_p} \frac{\partial E_y}{\partial y} = \frac{4\pi r_p}{A} c^2 \frac{d_i}{\left[ 1 + \frac{\sigma_y}{\sigma_x} \right]} \quad (54)$$

where  $d_i$  is the ion density.

Applying again the stability criterion for one period, we get the condition for trapping.

$$\frac{1}{A} [ A_c - k d_i ] < 1 \quad (55)$$

$$\text{with } k = \frac{\pi r_p s_B^2}{1 + \frac{\sigma_y}{\sigma_x}}$$

For zero initial density only masses  $A > A_c$  can accumulate. But if ions have already been accumulated, i.e.  $d_i \neq 0$ , the stability threshold decreases:

$$A > A_c - k d_i \quad (56)$$

The accumulation limit for a bunched beam is reached, if the focusing space charge force of the bunch corresponds to the integrated defocusing force of the already accumulated ions in the drift space between two bunches. This is the case, if the total number of trapped ions corresponds to the total number of electrons in the machine.  
The maximum ion density is then given by:

$$d_i = \frac{A_c}{k} = \frac{N}{2\pi R} \cdot \frac{\beta}{2\pi \sigma_x \sigma_y} \quad (57)$$

defining the neutralization factor as:

$$\eta = \frac{d_i}{d_p} \quad (58)$$

where  $d_p$  is the particle density in the bunch, the maximum value is given by:

$$\eta_{max} = \frac{n_B l_B}{2\pi R} \quad (59)$$

with  $n_B$ , the number of bunches,  $l_B$  the bunch length and  $R$  the average machine radius.

For the local pressure increase, we then find:

$$p = p_0 \left( 1 + \eta \frac{d_p}{d_m} \right) \quad (60)$$

with the density of molecules in the residual gas (at  $T = 300$  °K):

$$d_m = 3.2 \cdot 10^{13} p_0 [\text{nTorr}] \quad (61)$$

Especially for small emittances the pressure increase could be large. On the other hand low emittance beams generate a strong over focusing for the ion, which leads to a higher critical mass. For a more complete picture of the pressure increase second ionization should be taken into account.

Another potentially harmful effect comes from a tune spread and tune shift generated by the linear ion space charge force, which is given by:

$$\Delta Q_y = \frac{r_p}{\gamma} \int \frac{d_i \beta_y}{1 + \left( \frac{\sigma_y}{\sigma_x} \right)} ds \quad (62)$$

( $\Delta Q_y$  is larger than  $\Delta Q_x$  for flat beams).

## 5. LIFETIME DUE TO TOUSCHEK SCATTERING [15-17]

### 5.1 Cross Section

Particles inside a bunch perform transverse betatron oscillations around the closed orbit. Due to a scattering effect two particles can transform their transverse momenta into longitudinal momenta.

If the new longitudinal momenta of the two particles are outside the momentum acceptance, the particles are lost. This effect was first recognized by Bruno Touschek at the ADA storage ring of Frascati. The differential cross section for Coulomb scattering of two particles with equal but opposite momenta in the non-relativistic approximation is given by the Möller formula:

$$\frac{d\sigma}{d\Omega} = \frac{4r_0^2}{(v/c)^2} \left[ \frac{4}{\sin^4 \theta} - \frac{3}{\sin^2 \theta} \right] \quad (63)$$

where:  $r_0$  is the classical electron radius;  
 $v$  is the relative velocity in the centre of mass system (CMS);  
 $\theta$  is the scattering angle.

Figure 8 shows the geometry of the scattering process.

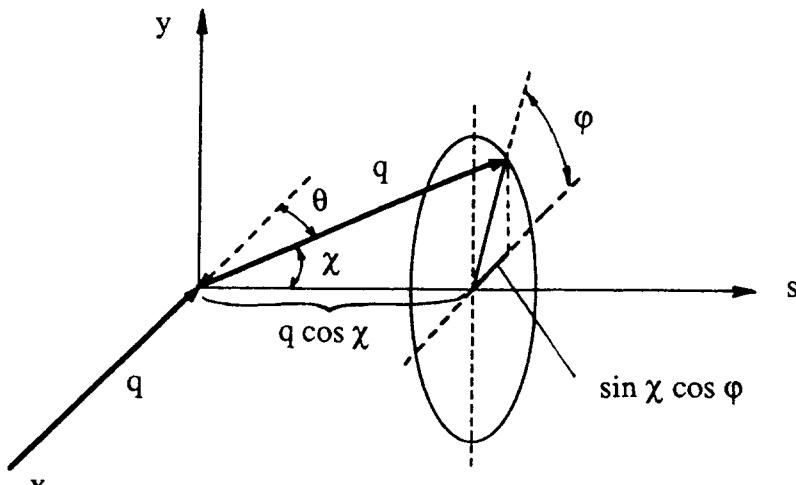


Fig. 8 Geometry of the process for Touschek scattering.

Following the original derivation by Bruck, we can introduce the dimensionless momentum:

$$q = \frac{p_x}{m_0 c} = \frac{1}{2} \left( \frac{v}{c} \right) \quad (64)$$

where  $p_x$  is the horizontal momentum of one particle and  $v$  the relative velocity in the CMS.

For the moment we consider a flat beam where the major collision effects happen in the horizontal plane. After elastic collision of the two particles in the CMS, the components of the particle momenta along the longitudinal direction are  $q(m_0 c) \cos\chi$ . Transferred back in the laboratory system, by

$$p_s = \gamma \left[ p'_s + \frac{\beta}{c} E' \right] \approx \gamma p_s$$

Since we assumed a non-relativistic motion in the CMS, we can neglect the second term in the bracket and get:

$$q(m_0 c) \gamma \cos\chi$$

The particle is lost if the component along the longitudinal direction is larger than the momentum acceptance, i.e.:

$$\begin{aligned} |q \cos\chi| &> \frac{\Delta p_{RF}}{\gamma} \quad \rightarrow \\ |\cos\chi| &> \frac{\Delta p_{RF}}{|q| \gamma} = \mu \end{aligned} \quad (65)$$

where  $\Delta p_{RF}$  ( $m_0 c$ ) is the maximum momentum accepted by the RF.

Using the geometric relation from figure 9, i.e.:

$$\cos\theta = \sin\chi \cos\phi$$

and

$$d\Omega = \sin\chi d\chi d\phi$$

we get for the total cross section:

$$\sigma = \frac{4r_0^2}{(v/c)^2} \int_0^{\arccos\mu} \sin\chi d\chi \int_{-\pi}^{\pi} d\phi \left[ \frac{4}{(1 - \sin^2\chi \cos^2\phi)^2} - \frac{3}{(1 - \sin^2\chi \cos^2\phi)} \right] \quad (66)$$

and finally after performing the integration:

$$\sigma = \frac{8\pi r_0^2}{(v/c)^4} \left[ \frac{1}{\mu^2} - 1 + \ln\mu \right] \quad (67)$$

(in the CMS)

## 5.2 Loss Rate and Lifetime

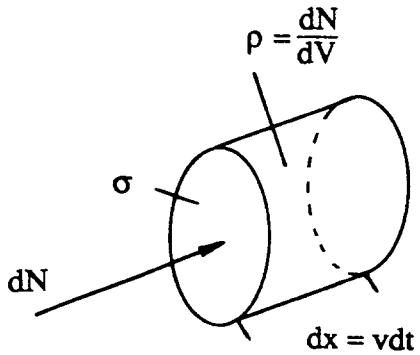


Fig. 9 Sketch for the derivation of the differential loss rate.

The differential loss rate of the particles, as sketched in figure 9, is given by:

$$d\left(\frac{dN}{dt}\right) = (\sigma v) \rho dN \quad (68)$$

Since incoming and scattered particles belong to the same ensemble, i.e.  $dN = \rho dV$ , we have:

$$d\left(\frac{dN}{dt}\right) = (\sigma v) \rho^2 dV \quad (69)$$

For the total loss rate we get:

$$\frac{dN}{dt} = \bar{\sigma v} \int p^2 dV \quad (\text{in the CMS}) \quad (70)$$

where we have assumed the approximation, that the product  $(\sigma v)$  is independent of the particle coordinates.

The integral can easily be performed in the Laboratory System (LS), where the distribution function is given by:

$$\rho(x, y, s) = \frac{N}{(2\pi)^{3/2} \sigma_x \sigma_y \sigma_s} e^{-\left[\frac{x^2}{2\sigma_x^2} + \frac{y^2}{2\sigma_y^2} + \frac{s^2}{2\sigma_s^2}\right]} \quad (71)$$

assuming decoupled Gaussian distributions for all directions  $(x, y, s)$ .

Performing the integration we get:

$$\int \rho^2 dV = \frac{N^2}{8\pi^{3/2} \sigma_x \sigma_s \sigma_y} \quad (72)$$

The average  $\bar{\sigma v}$  has to be done with regard to the change in relative momentum which leads to particle loss.

$$\overline{\sigma_v} = 2 \frac{\int_{\frac{\Delta p_{RF}}{\gamma}}^x g(q)(\sigma_v) dq}{\gamma} \quad (73)$$

where the factor 2 takes into account that two particles are lost in each scattering process.

The distribution function can be written as:

$$g(q) = \frac{1}{\sqrt{2\pi} \sigma_q} e^{-\frac{q^2}{2\sigma_q^2}} \quad \text{with } \sigma_q = \frac{\gamma \sigma_x'}{\sqrt{2}}$$

and substituting relations (70), (72), (73) and the bunch volume  $V_B = (4\pi)^{3/2} \sigma_x \sigma_y \sigma_z$

in  $\frac{1}{\tau} = \frac{1}{N} \frac{dN}{dt}$  we get:

$$\frac{1}{\tau} = \frac{\sqrt{\pi} r_0^2 c N_B (m_0 c)^3}{\gamma^2 V_B \sigma_{px}} * \int \frac{1}{q^3} \left[ \left( \frac{\gamma_q}{\Delta p_{RF}} \right)^2 - 1 - \ln \left( \frac{\gamma_q}{\Delta p_{RF}} \right) \right] e^{-\frac{q^2}{\sigma_{px}^2}} dq$$

After introducing the quantities:

$$u = \left( \frac{q}{\sigma_{px}} \right)^2 \quad \text{and} \quad \epsilon = \left( \frac{\Delta p_{RF}}{\gamma \sigma_{px}} \right)^2$$

we get the final form of the Touschek growth rate for a flat beam:

$$\frac{1}{\tau} = \frac{\sqrt{\pi} r_0^2 c}{\gamma^3} * \frac{N_B}{V_B \sigma_x'} * \frac{1}{\left( \frac{\Delta p_{RF}}{p_0} \right)^2} \left\{ \epsilon \int_{\epsilon}^{\infty} \frac{1}{u^2} \left[ \left( \frac{u}{\epsilon} \right) - \frac{1}{2} \ln \left( \frac{u}{\epsilon} \right) - 1 \right] e^{-u} du \right\} \quad (74)$$

where the expression in the curly bracket is usually abbreviated by  $C(\epsilon)$ , which for  $\epsilon < 1$  can be approximated by:

$$C(\epsilon) = \ln \left( \frac{1}{1.78 \epsilon} \right) - 1.5 \quad (75)$$

In a similar way we can derive the Touschek lifetime for a round beam [18]:

$$\frac{1}{\tau} = \frac{2\pi r_0^2 c}{\gamma^4} \frac{N_B}{V_B \sigma_x' \sigma_y'} \cdot \frac{D(\epsilon)}{\left( \frac{\Delta p_{RF}}{p_0} \right)} \quad (76)$$

with:

$$D(\epsilon) = \sqrt{\epsilon} \int_0^\infty \frac{1}{u^{3/2}} \left[ \left( \frac{u}{\epsilon} \right) - \frac{1}{2} \ln \left( \frac{u}{\epsilon} \right) - 1 \right] e^{-u} du \quad (77)$$

## 6. INTRABEAM SCATTERING [19-21]

Intrabeam scattering is usually used as a synonym for multiple small-angle Coulomb (or also Touschek) scattering inside the beam. This effect has first been completely described by A. Piwinski for a weak focusing machine. Later on it has been extended to a strong focusing machine by Mttingwa and Bjorken.

The derivation is only conceptually treated here, with the kinematics derived for the horizontal plane only.

### 6.1 Kinematics

The treatment given here closely follows that given in reference [20], but is for reasons of simplicity restricted to two dimensions only.

#### 6.1.1 Coordinate systems

Figure 10 shows the momenta of two colliding particles,  $\vec{p}_1$  and  $\vec{p}_2$  in the laboratory system.

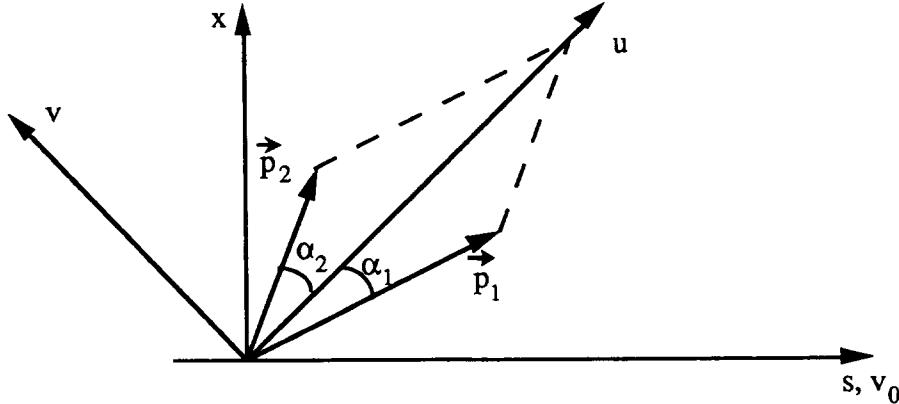


Fig. 10 Momenta of scattered particles in the laboratory system.

The beam is moving in direction "s", with velocity " $v_0$ ". The momenta in this coordinate system are given by:

$$\begin{aligned} \vec{p}_i &= p_i \left( \cos [\operatorname{arctg} (\dot{x}_i)], \sin [\operatorname{arctg} (\dot{x}_i)] \right) \\ &\approx p_i (1, \dot{x}_i) \end{aligned} \quad (78)$$

where  $x_i'$  is the slope of the particle trajectory in the x-plane and the index  $i = 1,2$  refers to the two particles.

In the (u,v)-coordinate system the components for the momenta are:

$$\begin{aligned} p_{1u} &= p_1 \cos \alpha_1, & p_{2u} &= p_2 \cos \alpha_2 \\ p_{1v} &= -p_1 \sin \alpha_1, & p_{2v} &= p_2 \sin \alpha_2 \end{aligned}$$

### 6.1.2 Transformation to the centre-of-mass system (CMS)

To treat the collision in a non-relativistic way, we now perform a transformation to the CMS (where the quantities are marked by primes). The relevant set of transformations is given by:

$$\begin{aligned} E_i' &= \gamma_u (E_i - \beta_u c p_{iu}) \\ p_{iu}' &= \gamma_u \left( p_{iu} - \frac{\beta_u}{c} E_i \right) \\ p_{iv}' &= p_{iv} \\ (i &= 1,2) \end{aligned} \tag{79}$$

$\beta_u$  and  $\gamma_u$  are then fixed by the conditions:

a)  
with

$$\begin{aligned} p_{1u}' + p_{2u}' &= 0 \\ p_i &= \frac{E_0}{c} \beta_i p_i \\ E_i &= E_0 p_i \end{aligned}$$

we get:

$$\beta_u = \frac{\beta_1 \gamma_1 \cos \alpha_1 + \beta_2 \gamma_2 \cos \alpha_2}{\gamma_1 + \gamma_2}$$

b)  
or

$$\begin{aligned} p_{1v}' + p_{2v}' &= 0 \\ \beta_1 \gamma_1 \sin \alpha_1 &= \beta_2 \gamma_2 \sin \alpha_2, \end{aligned}$$

### 6.1.3 Change of momenta in the CMS

In the CMS the two particles are deviated in their direction of motion by the angle  $\theta'$ , as indicated in figure 11.

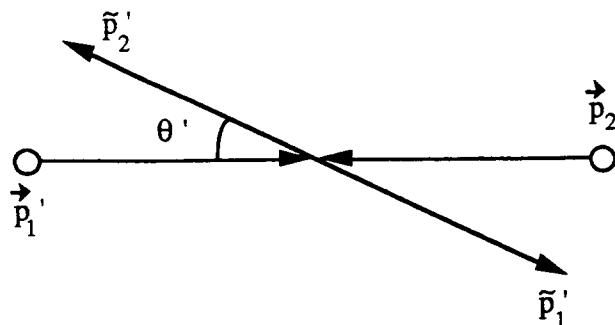


Fig. 11 Scattering process in the center-of-mass system.

The components of the new momenta in the CMS can be derived by the rotation:

$$\begin{bmatrix} \tilde{p}_u' \\ \tilde{p}_v' \end{bmatrix} = \begin{bmatrix} \cos\theta' & \sin\theta' \\ -\sin\theta' & \cos\theta' \end{bmatrix} \begin{bmatrix} p_u \\ p_v \end{bmatrix} = \begin{bmatrix} p_u \cos\theta' + p_v \sin\theta' \\ -p_u \sin\theta' + p_v \cos\theta' \end{bmatrix} \quad (80)$$

#### 6.1.4 Back transformation to the LS

The new momenta are now transformed back to the laboratory system:

$$\begin{aligned} \tilde{p}_{iu}' &= \gamma_u \left( \tilde{p}_{iu} + \frac{\beta_u}{c} E_i' \right) \\ \tilde{p}_{iv}' &= \tilde{p}_{iv} \end{aligned} \quad (81)$$

Since  $E_i' = E_{0i} \gamma_i$ , and the motion has been considered to be non-relativistic in the CMS, we can approximate:

$$\tilde{p}_{iu} = \gamma_u \tilde{p}_{iu} \quad (i = 1,2) \quad (82)$$

We can now derive the change in momenta due to the scattering process:

$$\bar{\delta}p_i = \tilde{p}_i - \bar{p}_i \quad (\text{in the } u,v \text{- system}) \quad (83)$$

Expressed in components, we find:

$$\begin{aligned} \delta p_{1u} &= p_1 \gamma_u \left[ \gamma_u \left( \cos\alpha_1 - \frac{\beta_u}{\beta_1} \right) (\cos\theta' - 1) - \sin\alpha_1 \sin\theta' \right] \\ \delta p_{1v} &= -p_1 \left[ \gamma_u \left( \cos\alpha_1 - \frac{\beta_u}{\beta_1} \right) \sin\theta' + \sin\alpha_1 (\cos\theta' - 1) \right] \\ \delta p_{2u} &= p_2 \gamma_u \left[ \gamma_u \left( \cos\alpha_2 - \frac{\beta_u}{\beta_2} \right) (\cos\theta' - 1) + \sin\alpha_2 \sin\theta' \right] \\ \delta p_{2v} &= -p_2 \left[ \gamma_u \left( \cos\alpha_2 - \frac{\beta_u}{\beta_2} \right) \sin\theta' - \sin\alpha_2 (\cos\theta' - 1) \right] \end{aligned} \quad (84)$$

## 6.2 Change of Equilibrium Beam Distribution

We now examine, how the change in momenta influences the equilibrium beam parameters, such as emittance and momentum spread.

### 6.2.1 Emittance

The emittance of a particle with betatron coordinates  $(x, x')$  at the position with optical parameters  $(\beta_x, \alpha_x)$  is given by:

$$\epsilon_x \beta_x = x^2 + (x\alpha_x + x'\beta_x)^2 \quad (85)$$

Due to the scattering process the coordinates  $(x, x')$  are changed. The displacement is changed, because the corresponding closed orbit is modified, i.e.:

$$\delta x = - \frac{\delta p}{p} D_x \quad (86)$$

Correspondingly the derivative of the dispersion changes the slope of the betatron coordinate. But in addition there is a modification of the transverse momentum which also causes a change of the slope, i.e.:

$$\delta x' = - \frac{\delta p}{p} D_x' + \frac{\delta p_x}{p_0} \quad (87)$$

For simplicity, we just consider the case  $\alpha_x = 0$  and  $D_x' = 0$ . We then find for the change in emittance:

$$\begin{aligned} \delta \varepsilon_x \beta_x &= [2x\delta_x + \delta_x^2] + \beta_x^2 [2x' \delta_{x'} + \delta_{x'}^2] \\ &= \left[ -2x D_x \frac{\delta p}{p} + D_x^2 \left( \frac{\delta p}{p} \right)^2 \right] + \beta_x^2 \left[ 2x' \frac{\delta p_x}{p_0} + \left( \frac{\delta p_x}{p_0} \right)^2 \right] \end{aligned} \quad (88)$$

### 6.2.2 Momentum spread

The invariant of the momentum spread for a bunched beam is given by:

$$H = \eta^2 + \frac{\dot{\eta}^2}{\Omega_s^2} \quad (89)$$

with  $\eta$  the relative momentum deviation and  $\Omega_s$  the synchrotron frequency. If the total momentum is changed, we find for the variation of the invariant.

$$\delta H = 2\eta \frac{\delta p}{p} + \left( \frac{\delta p}{p} \right)^2 \quad (90)$$

### 6.3 Statistical Averaging of the Invariants

Only the method will be sketched here, how the final results have been achieved can be found in reference [20].

First the invariants have to be averaged with respect to all scattering angles. Since the variations of the invariants are expressed by the changes in momenta, we have to find:

$$\langle \bar{\delta} p \rangle = \int_{\theta_m}^{\pi} \int_0^{2\pi} \bar{\delta} p \, d\sigma' \quad (91)$$

$$\langle \delta p^2 \rangle = \int_{\theta_m}^{\pi} \int_0^{2\pi} \delta p^2 \, d\sigma' \quad (92)$$

The lower limit of the first integral,  $\theta_m$  is given by a scattering angle, which corresponds to an impact parameter of half the beam size.

The differential of the cross section in the CMS is given by the non-relativistic Möller cross section, i.e.:

$$d\sigma' = \frac{p_0^2}{16 \beta^4} \frac{\sin \theta' \, d\theta' \, d\phi'}{\sin^4 \frac{\theta'}{2}} \quad (93)$$

which has been approximated for small scattering angles.

In the second step the mean value of the invariants of one particle can be calculated by averaging over all possible coordinates of the second particle. To derive the mean values for all particles, one has to additionally average in a third step over all possible coordinates of the first particle. Combining the second and third steps, all invariants have to be integrated with the distribution function.

$$P = f_s^2(s) f_z^2(z) f_x(x_1) f_\eta(\eta_1) f_{x'}(x_1') f_{z'}(z_1'). \\ f_x(x_2) f_\eta(\eta_2) f_{x'}(x_2') f_z(z_2') \quad (94)$$

The following relations hold between the particle coordinates:

$$\begin{aligned} s_1 &= s_2 \\ z_1 &= z_2 \\ x_1 + \eta_1 D_x &= x_2 \eta_2 D_x \end{aligned}$$

After having performed the averaging in the prescribed form, we can derive the invariant:

$$\langle H \rangle \left( \frac{1}{\gamma^2} - \alpha_c \right) + \frac{\langle \epsilon_x \rangle}{\beta_x} + \frac{\langle \epsilon_z \rangle}{\beta_z} = \text{const.} \quad (95)$$

There is a qualitatively different behaviour below and above transition. For  $\gamma < \gamma_T$ , i.e.  $\frac{1}{\gamma^2} > \alpha_c = \frac{1}{\gamma_T^2}$ , the motion is bounded.

An increase of one invariant can only happen at the expense of the other invariants. The particles behave like gas molecules in a closed box and an equilibrium of the distribution exists.

Above transition, i.e. for  $\frac{1}{\gamma^2} - \alpha_c < 0$ , the equilibrium quantities can increase beyond all limits and still fulfill the invariant condition.

## 7. LIFETIME DUE TO RESONANCE CROSSING [22]

Dangerous resonances are avoided by choosing an appropriate working point in the tune diagram. Due to various effects it might happen that the tune changes in time and resonances are crossed. One has to differ here between two effects depending on the speed of the tune variation.

### 7.1 Fast Crossing

For a sudden jump in tune within a time, which is small compared to the period a particle needs to complete its phase space cycle, the amplitude change is not very large. For the growth due to one rapid transversal of a one dimensional sum resonance we find:

$$\int_1^x \frac{dx}{x^{N-1}} = \frac{\Delta e}{N} \frac{\pi}{\sqrt{|\Delta Q_t|}} \quad (96)$$

where  $\Delta e$  is the resonance width,  $\Delta Q_t$  the tune change per turn and  $x$  the relative amplitude  $a_1/a_0$ . Depending on the order of the resonance,  $N$ , we get for the growth:

$$\left. \begin{array}{l} N=1 : x-1 \\ N=2 : \ln x \\ N>2 : \frac{1}{N} \left[ 1 - \frac{1}{x^N} \right] \end{array} \right\} = \frac{\Delta e}{N} \frac{\pi}{\sqrt{|\Delta Q_t|}} \quad (97)$$

For several random crossings the average of the amplitude growth has to be multiplied with the square root of the crossings.

For a sinusoidal variation of the time, the resonance will get an infinite number of sidebands spaced by  $\frac{Q_s}{N}$ , where  $Q_s$  is the synchrotron tune and  $N$  the order of the resonance. The lower order of these sidebands have usually to be avoided.

## 7.2 Slow Tune Variation

In the case of an adiabatic variation of the tune, the particles are trapped in the resonance island which is moving outwards and growing in phase space until the particle is lost at the aperture limit. Such an adiabatic variation of the tune can be generated for instance by a momentum diffusion (given by intrabeam scattering or beam gas scattering). If the diffusion constant is:

$$D_p = \frac{d}{dt} \left( \frac{\Delta p}{p} \right)_{rms}^2 \quad (98)$$

we find a related diffusion of the tune (given for instance by the non-compensated terms of the chromaticity) with the diffusion constant:

$$D_Q = (Q') D_p \quad (99)$$

The fractional loss rate at time  $t$  is given by:

$$\frac{1}{N} \frac{dN}{dt} = \frac{2p_T}{\Delta Q} \sqrt{\frac{D_Q}{\pi t}} e^{-\frac{8Q^2}{4D_Q} t} \quad (100)$$

with  $p_T$  the trapping probability, which depends on the resonance strength, the particle amplitude and the rate of change.

## 8 LIFETIME MEASUREMENT

The measured particle loss rate in an accelerator is given by the sum over all contributions, i.e.:

$$\frac{1}{\tau} = \sum_i \frac{1}{\tau_i} \quad (101)$$

It is not always easy to derive from measurements, how the individual effects contribute to the total lifetime. The difference in dependence on various machine parameters (such as energy, gas pressure, beam size, aperture, etc.) must be utilized to distinguish between the various effects. This also requires a good knowledge of the vacuum pressure, which has to be derived from the monitored pressure. If vacuum gauges are placed near the vacuum pumps, the monitored average pressure could be substantially lower than the real one. A simulation can be performed to derive the average pressure, which again is difficult, since reliable information on the pumping speed and details on desorption effects are needed.

For the interpretation of lifetime measurements, the difference in dependence of the various lifetime effects on machine parameters such as energy, beam size, aperture, etc. has to be utilized.

## ACKNOWLEDGEMENTS

The author thanks C.J. Bocchetta for reading the manuscript and making suggestions for improvements.

\* \* \*

## REFERENCES

- [1] A.A. Kolomensky and A.N. Lebedev, Theory of Cyclic Accelerators, North-Holland, 1966.
- [2] H. Risken, The Fokker-Planck Equation, Springer Verlag, Berlin, 1984.
- [3] M. Sands, The Physics of Electron Storage Rings, SLAC Report No. 121, 1970.
- [4] A. Piwinski, Beam Losses and Lifetime, Proc. of the CERN Accelerator School, CERN 85-19, 1985.
- [5] A. Chao, Quantum Lifetime in Electron Storage Rings, SLAC-PUB-1884, 1977.
- [6] D.A. Edwards and M.J. Syphers, An Introduction to the Physics of Particle Accelerators, Fermi National Accelerator Laboratory.
- [7] J. Le Duff, Current and Density Limitations in Existing Electron Storage Rings, Nucl. Instr. Methods, A239, 83, 1985.
- [8] J.D. Jackson, Classical Electrodynamics, John Wiley and Sons, New York, 1975.
- [9] P.H. Kung, The Bremsstrahlung Lifetime in a Storage Ring, SSRL ACD-Note 88, 1990.
- [10] H.A. Bethe and W. Heitler, Proc.R.Soc., Vol. A146, 83, 1934.
- [11] J.A. Wheeler and W.E. Lamb, Phys.Rev., Vol. 55, 858, 1939.
- [12] R.D. Kohaupt, Mechanismus der Ionenabsaugung im Electron Positron Speicherring DORIS, DESY h1-71/2, 1971.
- [13] Y. Baconnier, Neutralization of Accelerator Beams by Ionization of the Residual Gas, CERN/PS/PSR 84-24, 1984.
- [14] G. Brianti, The Stability of Ions in Bunched-Beam Machines, Proc. of the CERN Accelerator School, CERN 84-15, 1984.
- [15] C. Bernardini, B. Touschek et al, Phys. Rev. Lett. 10, 1963.
- [16] H. Bruck, Accélérateurs Circulaires de Particles, Presses Universitaires de France, Paris, 1966.
- [17] J. Le Duff, Single and Multiple Touschek Effects, Proc. of the CERN Accelerator School, CERN 89-01, 1989.

- [18] Y. Miyahara, A New Formula for the Lifetime of a Round Beam Caused by the Touschek Effect in an Electron Storage Ring, Japanese Journal of Appl.Phys., Vol.24, No.9, 1985.
- [19] H. Bruck and J. LeDuff, Beam Enlargement in Storage Rings by Multiple Coulomb Scattering, Proc. 5th Int.Conf. on High Energy Accel., Frascati, 1965.
- [20] A. Piwinski, Proc. 9th Int. Conf. on High Energy Accel., 105, 1974.
- [21] J.D. Bjorken and S.K. Mtingwa, Particle Accelerator, Vol. 13, 115, 1983.
- [22] G. Guignard, A General Treatment of Resonances in Accelerators, CERN 78-11, 1978.



# SYNCHROTRON RADIATION

*R.P. Walker*  
Sincrotrone Trieste, Italy

## Abstract

The basic properties of synchrotron radiation are described, and their relevance to the design of electron and proton rings is discussed. The development of specialized sources of synchrotron radiation is also considered.

## 1. INTRODUCTION

The electromagnetic radiation emitted by a charged particle beam in a circular accelerator is termed "synchrotron radiation" (SR) after its first visual observation nearly 50 years ago in the General Electric (G.E.) 70 MeV Synchrotron. The theoretical basis for understanding synchrotron radiation however goes back much further. Maxwell's equations (1873) made it clear that changing charge densities would radiate electromagnetic waves, and Hertz demonstrated these waves in 1887. The first more directly relevant development was the publication of Liénard's paper entitled "The electric and magnetic field produced by an electric charge concentrated at a point and in arbitrary motion" [1]. This work includes the energy loss formula for particles travelling on a circular path with relativistic velocities. Later, Schott published a detailed essay that included also the angular and frequency distribution of the radiation, as well as the polarization properties [2].

No further interest appears to have been taken in the topic until the early 1940's, when theoretical work in the Soviet Union showed that the energy loss may pose a limit on the maximum energy obtainable in the betatron [3]. Then in 1946 Blewett measured the energy loss due to SR in the G.E. 100 MeV betatron and found agreement with theory, but failed to detect the radiation after searching in the microwave region of the spectrum [4]. It was later pointed out by Schwinger however that the spectrum peaks at a much higher harmonic of the orbit frequency and that the power in the microwave region is negligible [5]. The first direct observation of the radiation as a "small spot of brilliant white light" occurred by chance in the following year at the G.E. 70 MeV synchrotron, when a technician was looking into the transparent vacuum chamber [6]. Later accounts of the event may be found in Refs. [7,8].

Soon after the discovery of SR the first systematic studies of the spectral distribution of the radiation were carried out in the visible part of the spectrum [9]. The results showed agreement with theory, which had been re-derived by Schwinger and expressed in a convenient form for practical computation [10]. At the 300 MeV synchrotron at Cornell experiments were carried out which accurately confirmed that the energy loss rate was proportional to the fourth power of electron energy, in agreement with theory [11]. Later, detailed spectral measurements were carried out in the UV and soft X-ray region by Tomboulian and Hartman [12]. Agreement between theory and measurement was also reported in the work carried out at the synchrotron of the Lebedev Institute, Moscow [13].

The emission of synchrotron radiation exerts a strong influence on the electron beam dynamics. It was already known before the first observation of SR that the energy loss would lead to a damping of the energy or synchrotron oscillations, a process known as Radiation Damping [14–16]. It was subsequently discovered, firstly by means of a quantum mechanical treatment [17] and later by a classical approach [18], that quantum fluctuations in the emission would give rise to a growth of oscillation amplitude and that the combined effect of the two processes would lead to a stable equilibrium [18]. Soon after, the same processes were found to occur also for the betatron oscillations [19,20]. Experimental verification of these effects was carried out at the synchrotron of the Lebedev Institute, Moscow [21].

Synchrotron radiation is of major importance in the design of electron synchrotrons and storage rings. The energy loss strongly affects the size of accelerator needed to reach a given energy. The design of the magnet lattice affects the processes of radiation damping and quantum excitation, and determines whether an equilibrium can be reached and the magnitude of the resulting beam dimensions. The emission of SR also has important implications for the design of ring components. A powerful r.f. system is needed to replace the energy lost due to SR. The design of the vacuum system must take into account the heat generated and the large amount of gas desorbed when SR beams impinge on the vacuum chamber walls. At high energies it becomes necessary to shield the vacuum chamber to prevent high energy photons escaping into the air and causing damage to other accelerator components. On the positive side, SR is widely used as a sensitive means of observing the electron beam.

Synchrotron radiation is also of major importance as a source of radiation for experiments. Following the work at Cornell in the early 50's it was pointed out that SR would be a useful source both for absorption measurements and as a standard for calibrating detectors in the VUV (vacuum ultra-violet) region of the spectrum [12]. Subsequently, in the early 1960's, several SR facilities were set up on rings built initially for High Energy Physics, in the U.S.A., Japan and in Europe. So successful was this "parasitic" use of the SR that a second generation of storage rings were built, for dedicated use as synchrotron radiation sources. At present a third generation of more sophisticated rings are under construction, and further developments are likely to continue.

In this Chapter the basic properties of synchrotron radiation and the way it affects the design of electron rings are described. Subsequent Chapters consider in more detail the effect of SR on beam dynamics. The development of specialized SR sources is treated briefly in Section 6. SR is emitted by any accelerated charge, but until recently particle energies have been sufficiently low that the effects have needed to be considered only in the case of electrons. For the next generation of high energy proton machines however the effects of SR emission will become significant. Although the treatment in this and subsequent chapters will be directed to the most common case, i.e. electrons, the formulae presented are valid for both electrons and protons unless indicated otherwise. Some special aspects of SR emission in proton rings are considered in Section 7.

## 2. BASIC PROPERTIES

### 2.1 Radiated power

The instantaneous power radiated by a moving electron, valid in the relativistic case, is given by the Liénard formula:

$$P = \frac{2}{3} \frac{e^2}{4\pi\epsilon_0 c} \gamma^6 \left[ \dot{\beta}^2 - (\boldsymbol{\beta} \wedge \dot{\boldsymbol{\beta}})^2 \right] \quad (1)$$

where  $\beta c$  is the velocity and  $\gamma((1-\beta^2)^{-1/2})$  is the usual relativistic factor. The result is most easily derived by a Lorentz transformation of the result for a non-relativistic electron [22], obtained by Larmor:

$$P = \frac{2}{3} \frac{e^2}{4\pi\epsilon_0 c} \frac{\dot{v}^2}{c^2} \quad (2)$$

We consider first the case of *linear motion*, for which  $\boldsymbol{\beta} \wedge \dot{\boldsymbol{\beta}} = 0$ . In this case it can be shown that:

$$P = \frac{2}{3} \frac{e^2}{4\pi\epsilon_0 m^2 c^3} \left( \frac{dp}{dt} \right)^2 \quad (3)$$

where  $\mathbf{p}$  is the relativistic momentum ( $\gamma\beta mc$ ). Since the rate of change of momentum ( $d\mathbf{p}/dt$ ) is equal to the gain in energy per unit distance ( $dE/ds$ ), we can write the ratio of the rate of energy loss to the rate of energy gain as follows:

$$\frac{2}{3} \frac{e^2}{4\pi\epsilon_0 m^2 c^4} \left( \frac{dE}{ds} \right) = \frac{2}{3} \frac{r_0}{mc^2} \frac{dE}{ds} \quad (4)$$

Thus, unless the rate of energy gain is comparable to the rest mass energy (0.511 MeV) in a distance equal to the classical electron radius ( $2.8 \cdot 10^{-15}$  m) the radiative energy loss is negligible. As a further demonstration of this fact, it can be calculated that the energy loss in a typical 2 GeV linac, 200 m long, is only  $7 \cdot 10^{-5}$  eV.

In the case of *circular motion*, we have  $\mathbf{B} \wedge \mathbf{p} = \beta \mathbf{\beta}$ . In this case it may be shown that:

$$P = \frac{2}{3} \frac{e^2}{4\pi\epsilon_0 m^2 c^3} \gamma^2 \left( \frac{d\mathbf{p}}{dt} \right)^2 \quad (5)$$

Thus, for the same applied force ( $d\mathbf{p}/dt$ ) the power radiated is a factor of  $\gamma^2$  larger than for linear motion; moreover, it should be remembered that a given force is much more easily generated with a magnetic field compared to an electric field: a field of 1 Tesla produces the same force as an electric field of 300 MV/m.

For circular motion we have from the above that the instantaneous rate of power emitted is:

$$P = \frac{2}{3} \frac{e^2 c}{4\pi\epsilon_0} \frac{\beta^4 \gamma^4}{\rho^2} \quad (6)$$

Evaluating this expression we find that the resulting instantaneous power is apparently very small - only  $7 \mu\text{W}$  even in the case of LEP at 100 GeV. A more directly meaningful parameter is the total energy loss per turn:

$$U_0 = \oint \frac{P}{\beta c} ds = \frac{2}{3} \frac{e^2}{4\pi\epsilon_0} \beta^3 \gamma^4 \oint \frac{ds}{\rho^2} \quad (7)$$

For an isomagnetic lattice (uniform bending radius in the bending magnets) the result simplifies to:

$$U_0 = \frac{e^2}{3\epsilon_0} \frac{\beta^3 \gamma^4}{\rho} \quad (8)$$

In practical units, for electrons:

$$U_0 [\text{eV}] = 8.85 \cdot 10^4 \frac{E^4 [\text{GeV}]}{\rho [\text{m}]} = 2.65 \cdot 10^4 E^3 [\text{GeV}] B [\text{T}] \quad (9)$$

Since an accelerating voltage at least as great as the energy loss per turn is required for the electron to be stable it can be seen from the examples in Table 1 that this can be an appreciable quantity. For a beam of electrons the total power emitted can be obtained as follows:

$$P_b = \frac{U_0 N_e}{T_0} = \frac{U_0 I_b}{e} \quad (10)$$

where  $N_e$  is the number of electrons in the ring,  $T_0$  the orbit time, and  $I_b$  the average beam current. In practical units therefore the total power (in Watts) is simply the energy loss per turn (in eV) multiplied by the beam current (in Amps). With beam currents in the few mA to few 100 mA range it can be seen that the total SR power can be very large. This is also the power that the r.f. system must provide to make up for the losses due to SR emission.

It can be seen that there is a rapid increase in SR emission with energy, varying as  $E^4$ . There is therefore also a strong dependence on the mass of the particle,  $\sim 1/m^4$ , and so for protons with the same energy and bending radius the quantities are reduced by a large factor,  $\sim 10^{13}$ , compared to electrons.

**Table 1**  
Energy loss per turn ( $U_0$ ) and critical energy ( $\varepsilon_c$ ) in various electron storage rings

Ring	$E$ [GeV]	$\rho$ [m]	$U_0$ [MeV]	$\varepsilon_c$ [keV]
EPA	0.6	1.43	0.008	0.34
SRS	2.0	5.6	0.25	3.2
DORIS	5.0	12.3	4.5	22.5
PEP	18	166	56.1	78.1
HERA	30	550	140	109
LEP	55	3100	261	119
	100	3100	2855	715

## 2.2 Qualitative description of angular and spectral properties

Consider an electron travelling with speed  $\beta c$  along a circular path with constant radius  $\rho$  in a bending magnet (Fig. 1a). Viewed in a frame moving with the same speed in a direction tangent to the curve at some point  $P$ , the motion appears as shown in Fig. 1b. There is clearly an acceleration of the electron in the  $x'$  direction which gives rise to the emission of radiation. Since the motion in this frame is non-relativistic the radiation pattern is that of the familiar dipole radiation with  $\sin^2\theta$  distribution, where  $\theta$  is the angle with respect to the acceleration ( $x'$ ) axis [22]. Transformation of this distribution into the laboratory frame results in a distribution which is strongly peaked in the forward direction. This is evident from the fact that the angle at which zero emission occurs ( $x'$  axis,  $\theta = 0$ ) transforms to an angle  $1/\gamma$  in the laboratory frame (Fig. 1c), which can be a very small angle, e.g. 1/10 mrad for 5 GeV electrons.

The compression of the radiation into a narrow range of angles has an important consequence for the spectrum of the radiation. The radiation received in a given direction corresponds to only a small arc of the electron trajectory, of length  $2\rho/\gamma$  (see Fig. 2). The radiation pulse length is the difference between the time it takes for the electron to travel around the arc from  $A$  to  $B$  ( $t_e$ ) and the time for the photons to travel directly from  $A$  to  $B$  ( $t_v$ ). The radiation pulse length is then:

$$\Delta t = t_e - t_v = \frac{2\rho}{\gamma\beta c} - \frac{2\rho \sin(1/\gamma)}{c} \approx \frac{\rho}{\gamma^2 c} \quad (11)$$

The short time interval implies a wide frequency range with typical frequency given approximately by  $\omega \sim 1/\Delta t \sim \gamma^2 c/\rho$ . In terms of the electron revolution frequency ( $\omega_{\text{rev}} \approx c/\rho$ ) we see that  $\omega/\omega_{\text{rev}} \sim \gamma^2$ . The spectrum therefore extends to very much higher frequencies than the orbit frequency.

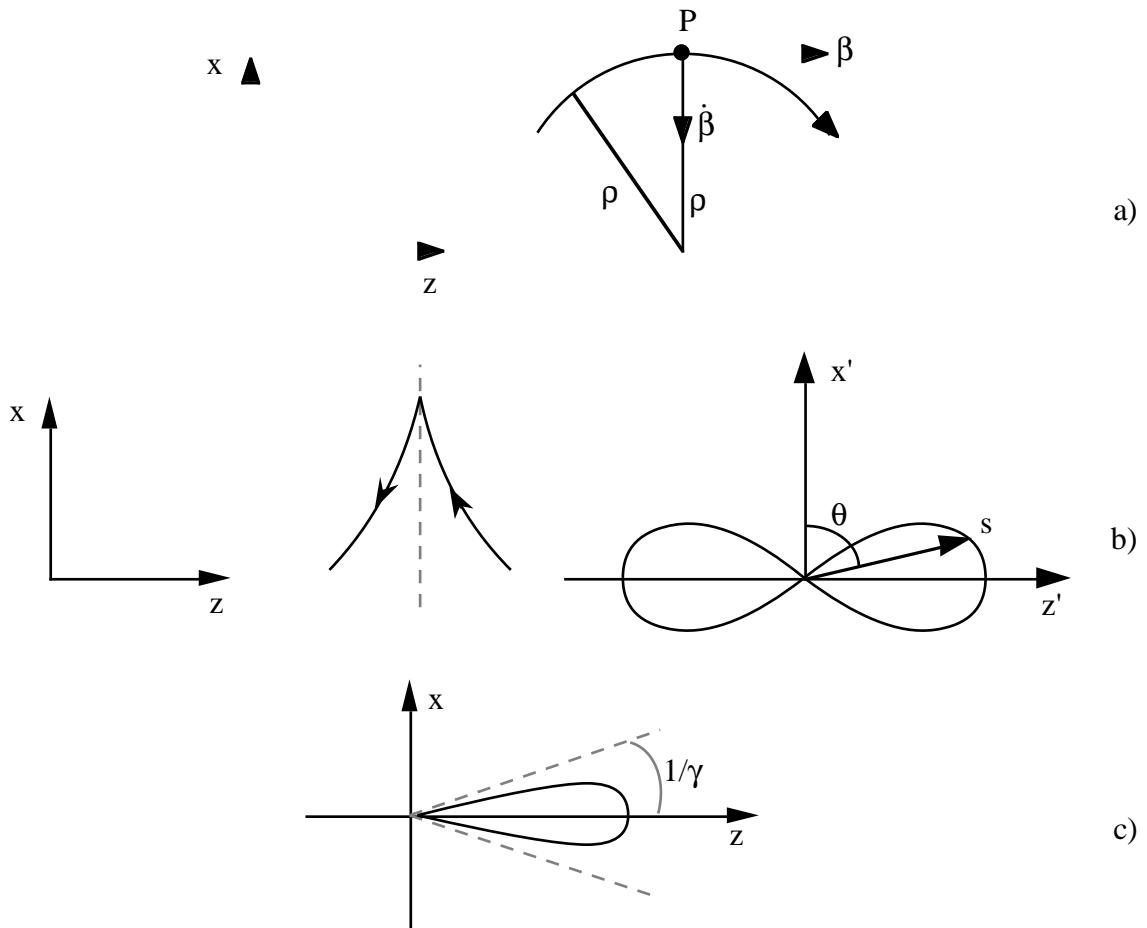


Fig. 1 Circular particle motion (a), the trajectory and dipole radiation emission pattern as seen in a frame moving with the average electron speed  $\beta c$  (b) and the corresponding radiation pattern transformed into the lab. frame (c).

As an example, if we take the parameters of the G.E. synchrotron ( $E = 70 \text{ MeV}$ ,  $\rho = 29.2 \text{ cm}$ ) we obtain a typical wavelength of about  $7000 \text{ \AA}$ , i.e. in the visible part of the spectrum. It is evident also that with higher electron energies,  $\geq 1 \text{ GeV}$ , the emission will extend into the X-ray region.

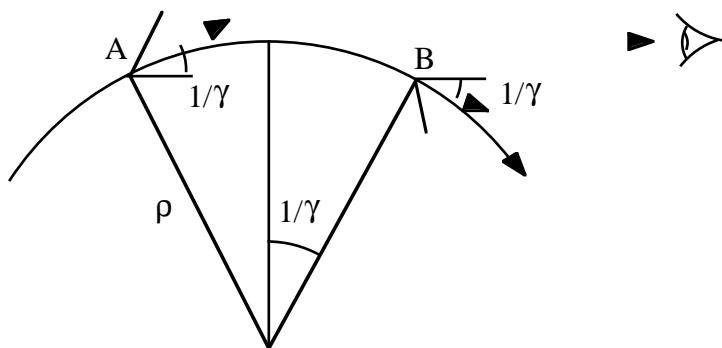


Fig. 2 Illustration of the limited arc of the trajectory which contributes to the radiation seen by an observer in a given direction

### 3. SPECTRAL AND ANGULAR PROPERTIES

A complete calculation of the spectral and angular distributions of the radiation is somewhat lengthy and therefore beyond the scope of this introduction and so here we only sketch the main features of the calculation and present the most important results. Further details may be obtained from various reports and textbooks [22–27].

#### 3.1 Observer and emitter time

For charges in relativistic motion a fundamental concept is the distinction between observer and emitter (or "retarded") time. An electron at position  $\mathbf{r}$  at time  $t'$  emits radiation towards an observer at a distance  $R(t')$  away in a direction defined by the unit vector  $\mathbf{n}(t')$ , see Fig. 3. The radiation arrives at the observer at a later time  $t$ , where:

$$t = t' + \frac{R(t')}{c} \quad (12)$$

Although this relation between the two times appears relatively simple, the fact that the distance  $R$  is changing with  $t'$  complicates the analysis significantly. Taking the derivative of the above it can be shown that:

$$\frac{dt}{dt'} = 1 - \mathbf{n} \cdot \beta = 1 - \beta \cos \theta \quad (13)$$

Thus, when the direction of motion ( $\beta$ ) is pointing close to the direction of the observer ( $\mathbf{n}$ ) we see that a time interval for the emitter ( $dt'$ ) is seen as a much shorter time interval ( $dt$ ) for the observer:

$$\frac{dt}{dt'} \approx \frac{1}{2} \left( \frac{1}{\gamma^2} + \theta^2 \right) \quad (14)$$

This is another description of the compression of the pulse length, with its consequent effect on the radiation spectrum, that was discussed in section 2.2. It also explains why the radiated intensity is very much greater for relativistic particles. The intensity of the radiation per unit solid angle at the observer can be related to the square of the *apparent* acceleration of the particle, i.e. the acceleration seen in the time frame of the observer, which is therefore increased by a large factor due to the time compression effect [27].

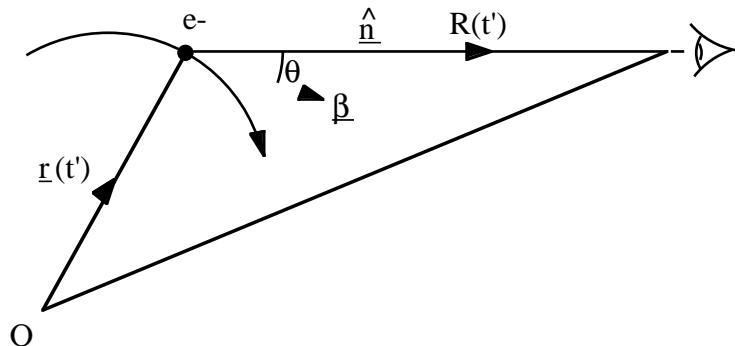


Fig. 3 Geometry of the emission of synchrotron radiation from a charged particle in arbitrary motion

### 3.2 Potentials and fields of a moving charge

The scalar electric and magnetic vector potentials of a slowly moving charge are given by the standard expressions:

$$V = \frac{e}{4\pi\epsilon_0 R} \quad A = \frac{e\mathbf{v}}{4\pi\epsilon_0 R} \quad (15)$$

where  $R$  is the distance from the charge to the observer and  $\mathbf{v}$  the charge velocity. In the case of a relativistically moving charge we must take into account that the potential created by the charge reaches the observer at a later time, according to Eq. (12). Charges moving towards the observer contribute for a longer time and therefore result in a larger potential [26]. It can be shown that the resulting expressions are similar, but includes the time correction factor Eq. (13):

$$V(t) = \frac{e}{4\pi\epsilon_0} \left[ \frac{1}{R(1 - \mathbf{n} \cdot \boldsymbol{\beta})} \right]_{ret.} \quad A(t) = \frac{e}{4\pi\epsilon_0} \left[ \frac{\mathbf{v}}{R(1 - \mathbf{n} \cdot \boldsymbol{\beta})} \right]_{ret.} \quad (16)$$

where  $[ ]_{ret.}$  means evaluated at the emitter, or retarded time  $t'$ .

The electric and magnetic fields can then be calculated by applying Maxwell's equations:

$$\mathbf{E} = -\nabla V - \mu_0 \frac{\partial \mathbf{A}}{\partial t} \quad \mathbf{B} = \mu_0 (\nabla \wedge \mathbf{A}) \quad (17)$$

The operation is not so straightforward however, since derivatives with respect to the observer's position involve changes in emitter time. The resulting expressions are given as follows:

$$\begin{aligned} \mathbf{E}(t) &= \frac{e}{4\pi\epsilon_0} \left[ \frac{\mathbf{n} - \boldsymbol{\beta}}{(1 - \mathbf{n} \cdot \boldsymbol{\beta})^3 R^2 \gamma^2} \right]_{ret.} + \frac{e}{4\pi\epsilon_0 c} \left[ \frac{\mathbf{n} \wedge \{(\mathbf{n} - \boldsymbol{\beta}) \wedge \dot{\boldsymbol{\beta}}\}}{(1 - \mathbf{n} \cdot \boldsymbol{\beta})^3 R} \right]_{ret.} \\ \mathbf{B}(t) &= \frac{\mathbf{n} \wedge \mathbf{E}}{c} \end{aligned} \quad (18)$$

From the above expressions we note the following features:

- $\mathbf{E}$  and  $\mathbf{B}$  are mutually perpendicular;
- for a static charge ( $\boldsymbol{\beta} = \dot{\boldsymbol{\beta}} = 0$ ) we recover Coulomb's Law  $\mathbf{E} = e \mathbf{n} / 4\pi\epsilon_0 R$ ;
- the first term is independent of acceleration and is proportional to  $1/R^2$ , and is therefore negligible at large distances; this is termed the "velocity field";
- the second term depends directly on the acceleration, and since it varies as  $1/R$  dominates at large distances; this is called the "acceleration field" or the "far field".

The second term is therefore the one that we shall consider. For this term we see that  $\mathbf{E}$  and  $\mathbf{B}$  are perpendicular to  $\mathbf{n}$ .

### 3.3 Power distribution

The flow of energy is described by the Poynting vector,  $\mathbf{S}$ :

$$\mathbf{S} = \frac{1}{\mu_0} (\mathbf{E} \wedge \mathbf{B}) \quad (19)$$

which is the energy passing through a unit area in direction  $\mathbf{S}$ . The energy received per unit solid angle in direction  $\mathbf{n}$  at distance  $R$  from the source is therefore:

$$\frac{dP}{d\Omega} = R^2 (\mathbf{S} \cdot \mathbf{n}) = \frac{1}{\mu_0 c} (R \mathbf{E})^2 \quad (20)$$

To express this in terms of the energy radiated per unit time of the emitter, we must include the factor in Eq. (13):

$$\frac{dP}{d\Omega}(t') = \frac{dP}{d\Omega} \frac{dt}{dt'} = \frac{dP}{d\Omega} (1 - \mathbf{n} \cdot \boldsymbol{\beta}) \quad (21)$$

Inserting the expression for the electric field, Eq. (18), gives the general result for the angular distribution of the instantaneous power radiated by the charge:

$$\frac{dP}{d\Omega}(t') = \frac{e^2}{(4\pi)^2 \epsilon_0 c} \frac{(\mathbf{n} \wedge \{(\mathbf{n} - \boldsymbol{\beta}) \wedge \dot{\boldsymbol{\beta}}\})^2}{(1 - \mathbf{n} \cdot \boldsymbol{\beta})^5} \quad (22)$$

In the case of circular motion ( $\mu = \mu_0 \cup \mu_J$ ) and in the relativistic case, the resulting distribution is given by:

$$\frac{dP}{d\Omega}(t') = \frac{e^2 c}{2\pi^2 \epsilon_0} \frac{\gamma^6}{\rho^2} \left( \frac{1 + 2\gamma^2 \theta^2 (1 - 2\cos^2 \phi) + \gamma^4 \theta^4}{(1 + \gamma^2 \theta^2)^5} \right) \quad (23)$$

(See Fig. 4). Thus, in agreement with the results of section 2, we see that the distribution is strongly peaked in the forward direction within a cone  $1/\gamma$ . In the non-relativistic limit the electric field, Eq. (18), reduces to:

$$\mathbf{E}(t) = \frac{e}{4\pi\epsilon_0 c} \frac{\mathbf{n} \wedge (\mathbf{n} \wedge \dot{\boldsymbol{\beta}})}{R} \quad (24)$$

and hence the power distribution becomes:

$$\frac{dP}{d\Omega} = \frac{e^2}{(4\pi)^2 \epsilon_0 c} (\mathbf{n} \wedge \{\mathbf{n} \wedge \dot{\boldsymbol{\beta}}\})^2 = \frac{e^2}{4\pi^2 \epsilon_0 c} \dot{\beta}^2 \sin^2 \Theta \quad (25)$$

with  $\Theta$  as the angle between the acceleration axis and the radiation direction. Figure 5 shows a comparison of the instantaneous power distribution in the two cases.

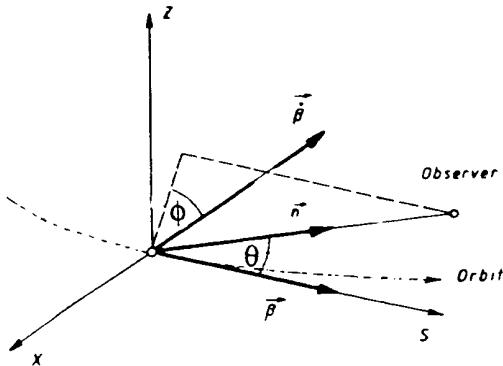


Fig. 4  
Definition of angles  $\theta$  and  $\phi$  in the instantaneous radiation emission, Eq.(23)

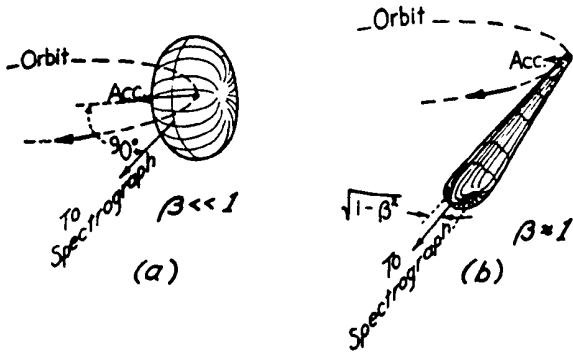


Fig. 5  
Illustration of the instantaneous radiation pattern from a charged particle in a circular orbit in the non-relativistic (a) and the relativistic cases (b), from Ref.[12]

### 3.4 Spectral and angular distributions

In order to obtain the frequency dependence we define the Fourier transform of the electric field,  $\tilde{E}(\omega)$ , and its inverse:

$$\tilde{E}(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} E(t) e^{i\omega t} dt \quad E(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \tilde{E}(\omega) e^{-i\omega t} d\omega \quad (26)$$

The total energy received per unit solid angle during one passage of the electron past the observer is from Eq. (20):

$$\frac{dW}{d\Omega} = \int \frac{dP}{d\Omega} dt = \frac{1}{\mu_0 c} \int_{-\infty}^{\infty} (RE)^2 dt \quad (27)$$

Inserting the expression for the electric field and simplifying, results in the following integral over frequency:

$$\frac{dW}{d\Omega} = \frac{1}{\mu_0 c} \int_0^{\infty} 2 |R\tilde{E}(\omega)|^2 d\omega \quad (28)$$

The integrand in the above is therefore the required spectral and angular density. Inserting now the expression for  $E(\omega)$  we obtain finally:

$$\frac{d^2 W}{d\omega d\Omega} = \frac{1}{2\pi\mu_0 c} \left| \int_{-\infty}^{\infty} (RE) e^{i\omega t} dt \right|^2 \quad (29)$$

Thus, the total energy received per unit solid angle per unit frequency interval is the modulus-squared of the Fourier transform of the electric field seen by the observer.

Inserting the expression for the Electric field, Eq. (18), we obtain:

$$\frac{d^2W}{d\Omega d\omega} = \frac{e^2}{16\pi^3 \epsilon_0 c} \left| \int_{-\infty}^{\infty} \left[ \frac{\mathbf{n} \wedge \{(\mathbf{n} - \boldsymbol{\beta}) \wedge \dot{\boldsymbol{\beta}}\}}{(1 - \mathbf{n} \cdot \boldsymbol{\beta})^3} \right]_{ret.} e^{i\omega t} dt \right|^2 \quad (30)$$

Thus, the general prescription for calculating the spectral/angular distribution for arbitrary electron motion is as follows: calculate the electron motion ( $\mathbf{r}$ ,  $\beta$ ,  $\dot{\beta}$ ) and the direction and distance to the observer ( $\mathbf{n}$ ,  $R$ ) as a function of retarded (emitter) time,  $t'$ ; calculate the electric field, Eq. (18); express as a function of observer time using  $t = t' + R(t')/c$  and perform a Fourier transform. Alternatively, the integration can be expressed directly in terms of emitter time  $t'$ :

$$\frac{d^2W}{d\Omega d\omega} = \frac{e^2}{16\pi^3 \epsilon_0 c} \left| \int_{-\infty}^{\infty} \frac{\mathbf{n} \wedge \{(\mathbf{n} - \boldsymbol{\beta}) \wedge \dot{\boldsymbol{\beta}}\}}{(1 - \mathbf{n} \cdot \boldsymbol{\beta})^2} e^{i\omega \left( t' + \frac{R(t')}{c} \right)} dt' \right|^2 \quad (31)$$

For circular motion the result can be derived analytically in terms of Airy integrals or alternatively modified Bessel functions:

$$\frac{d^2W}{d\Omega d\omega} = \frac{e^2}{16\pi^3 \epsilon_0 c} \gamma^2 \left( \frac{\omega}{\omega_c} \right)^2 (1 + \gamma^2 \psi^2)^2 \left[ K_{2/3}^2(\xi) + \frac{\gamma^2 \psi^2}{1 + \gamma^2 \psi^2} K_{1/3}^2(\xi) \right] \quad (32)$$

where,

$$\xi = \frac{\omega}{\omega_c} \frac{(1 + \gamma^2 \psi^2)^{3/2}}{2} \quad (33)$$

Because of the horizontal motion of the source the result depends only on the observation angle in the vertical plane,  $\psi$  (Fig. 6).  $\omega_c$  is the critical frequency defined by:

$$\omega_c = \frac{3}{2} \frac{\gamma^3 c}{\rho} \quad (34)$$

This is very similar to the value of the typical frequency derived earlier from consideration of the pulse length. Note that some authors use a value which is twice the above value [22]. Practical formulae for calculating the critical frequency are given in the Appendix.

The two terms in Eq. (32) above correspond to the radiation polarized in the horizontal ( $\sigma$ ) and vertical ( $\pi$ ) planes. The angular distributions at several different frequencies are shown in Fig. 7, for both polarization components, normalized in each case to the peak value on axis ( $\psi = 0$ ). On axis the radiation is linearly polarized in the horizontal plane. As  $\psi$  increases the vertically polarized component increases, and because of a  $\pi/2$  phase shift, not directly apparent from the intensities in Eq. (32) above, this results in circularly polarized radiation. The angular divergence changes markedly with the radiation frequency. At  $\omega = \omega_c$ , and approximating as a Gaussian function, the effective standard deviation is  $0.57/\gamma$ .

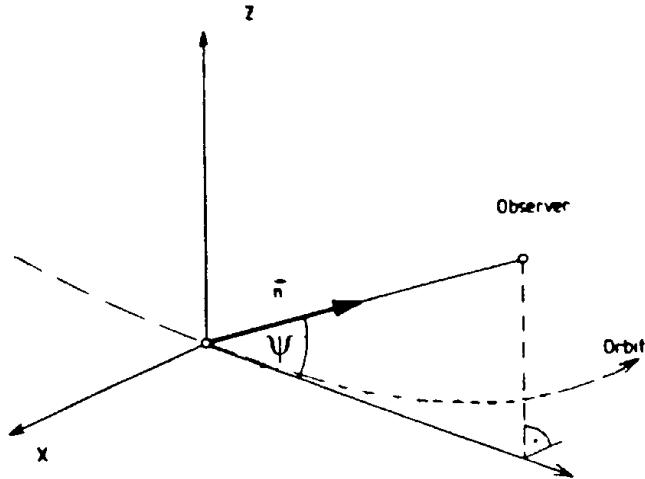


Fig. 6 Geometry of the emission of synchrotron radiation in the case of circular motion

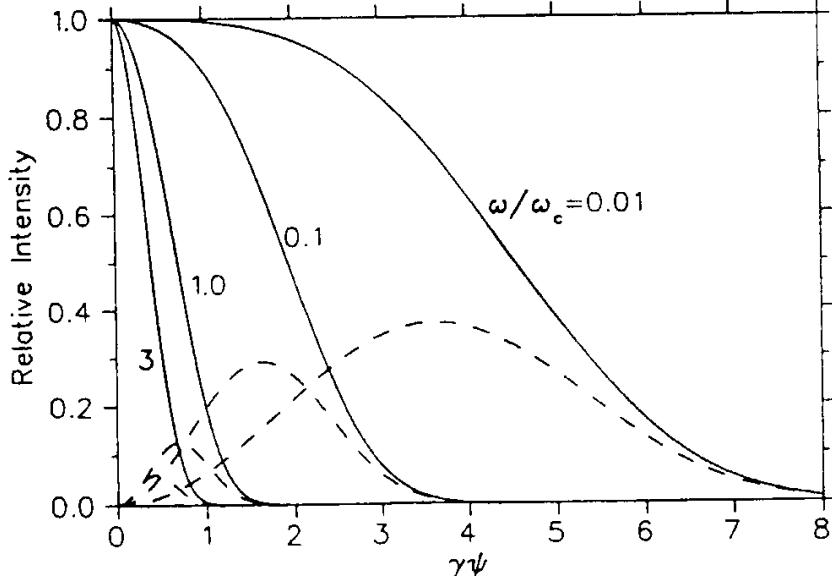


Fig. 7 Angular distribution of synchrotron radiation polarized in the orbit plane (solid lines) and in the plane perpendicular to the orbit plane (dotted lines)

The variation of the peak intensity on-axis as a function of frequency is given by:

$$\frac{d^2W}{d\Omega d\omega} = \frac{e^2}{16\pi^3 \epsilon_0 c} \gamma^2 \left(\frac{\omega}{\omega_c}\right)^2 K_{2/3}^2 \left(\frac{\omega}{2\omega_c}\right) = \frac{e^2}{16\pi^3 \epsilon_0 c} \gamma^2 H_2 \left(\frac{\omega}{\omega_c}\right) \quad (35)$$

The function  $H_2$  is shown in Fig. 8. As anticipated in section 2.2, it can be seen that the spectral range is very broad. The peak value occurs close to the critical frequency ( $\omega/\omega_c = 0.83$ ) where  $H_2$  has a value of 1.47.

Integrating Eq. (32) over all angles one obtains the energy radiated per unit frequency interval, per turn:

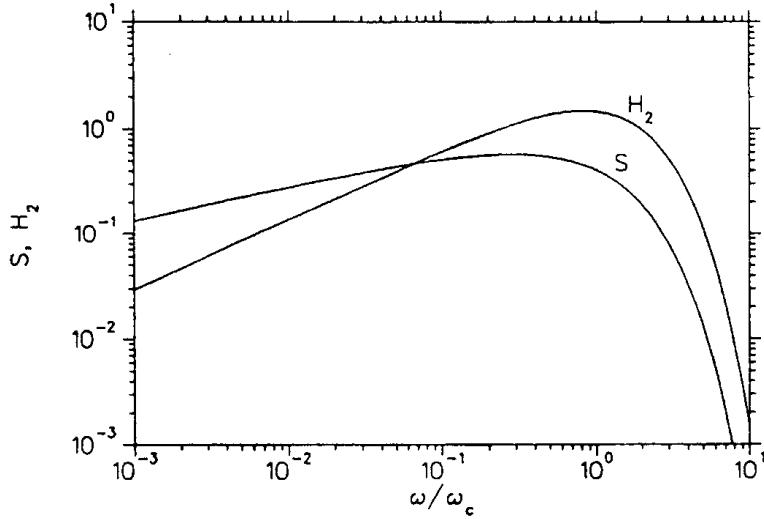


Fig. 8 Functions describing the spectral distributions of peak angular density ( $H_2$ ) and intensity integrated over vertical angle ( $S$ )

$$\frac{dW}{d\omega} = \frac{\sqrt{3} e^2}{4\pi\epsilon_0 c} \gamma(\omega / \omega_c) \int_{\omega/\omega_c}^{\infty} K_{5/3}(\omega / \omega_c) d(\omega / \omega_c) \quad (36)$$

This can be written as follows:

$$\frac{dW}{d\omega} = \frac{U_0}{\omega_c} S(\omega / \omega_c) \quad (37)$$

where  $S(\omega/\omega_c)$  is the normalized spectrum, shown in Fig. 8, defined by:

$$S(\omega / \omega_c) = \frac{9\sqrt{3}}{8\pi} \frac{\omega}{\omega_c} \int_{\omega/\omega_c}^{\infty} K_{5/3}(\omega / \omega_c) d(\omega / \omega_c) \quad (38)$$

The spectrum is normalized since:

$$\int_0^{\infty} S(\omega / \omega_c) d(\omega / \omega_c) = 1 \quad (39)$$

It should be noted also that:

$$\int_0^1 S(\omega / \omega_c) d(\omega / \omega_c) = 0.5 \quad (40)$$

In other words, half of the power is emitted above the critical frequency, and half below. The peak of the total spectrum occurs at  $\omega/\omega_c = 0.29$  at which point  $S = 0.57$ . The spectrum of vertically integrated intensity is also often written in terms of the function  $G_1(\omega/\omega_c)$  which includes all but the numerical factor in Eq. (38) and so the relation between the two is simply  $S(\omega/\omega_c) = 0.620 G_1(\omega/\omega_c)$ .

Integrating over frequency one obtains the total energy radiated per unit solid angle (per pass):

$$\frac{dW}{d\Omega} = \frac{7e^2}{64\pi\epsilon_0} \frac{\gamma^5}{\rho} \frac{1}{(1+\gamma^2\psi^2)^{5/2}} \left[ 1 + \frac{5\gamma^2\psi^2}{7(1+\gamma^2\psi^2)} \right] \quad (41)$$

This distribution is shown in Fig. 9.

The Appendix includes practical versions of the above formulae.

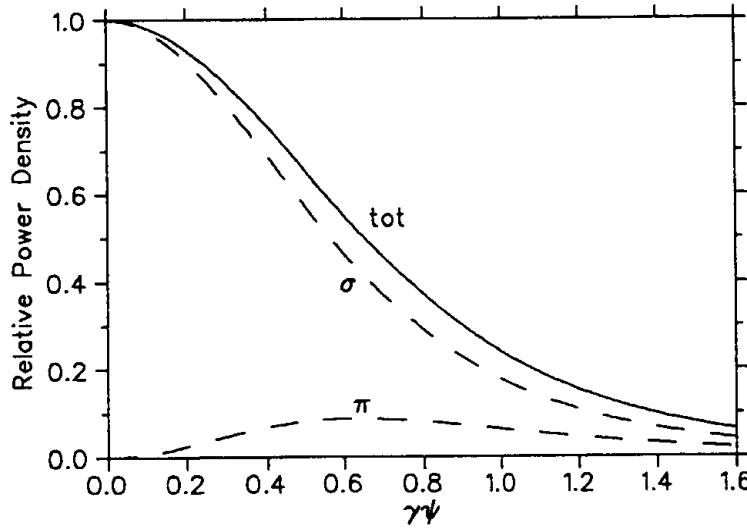


Fig. 9 Angular distribution of power density, polarized in the orbit plane ( $\sigma$ ), in the plane perpendicular to the orbit ( $\pi$ ), and total (tot).

#### 4. PHOTON DISTRIBUTION

So far we have considered that an electron radiates energy continuously, which is an acceptable assumption from the point of view of calculating the spectral properties etc. In reality however, radiation is emitted in discrete 'quanta' or photons, each with an energy,  $u$ , given by:

$$u = \hbar\omega \quad (42)$$

where  $\hbar$  = Planck's constant /  $2\pi$ . The distribution of the number of photons emitted as a function of angle and frequency is of interest for a number of reasons. One of the effects on beam dynamics ("quantum excitation") depends directly on the number of photons emitted as a function of energy. Also, since individual photons desorb gas molecules, the effect on the vacuum system depends on the number of photons, not simply total power. In addition, the common unit of intensity employed by SR experimenters is the spectral flux – the number of photons per second per unit frequency. Expressions for this quantity are given in the Appendix.

For the effect on beam dynamics the important quantity is the instantaneous rate of emission of photons. We define a quantity  $n(u)$ , the number of photons per second radiated (on average) by a single electron per unit energy interval:

$$n(u)\Delta u = \frac{dW}{d\omega} \frac{c}{2\pi\rho} \frac{1}{\hbar\omega} \Delta\omega \quad (43)$$

hence,

$$n(u) = \frac{P}{u_c^2} \frac{S(u/u_c)}{(u/u_c)} \quad (44)$$

where,

$$u_c = \hbar\omega_c \quad (45)$$

The photon distribution function is shown in Fig. 10.

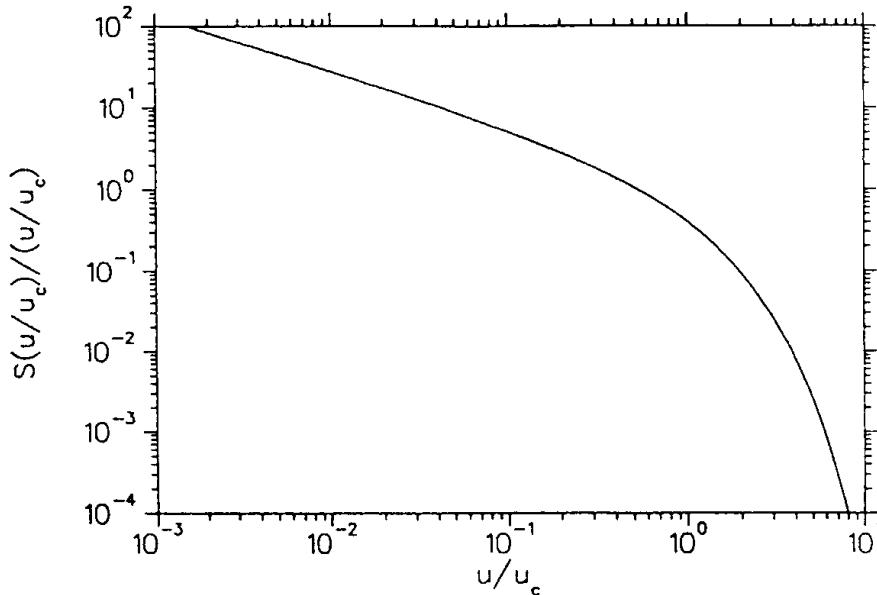


Fig. 10 Photon distribution function

The total number emitted (per electron per second) is then:

$$N = \int n(u) du = \frac{15\sqrt{3}}{8} \frac{P}{u_c} = \frac{5}{2\sqrt{3}} \frac{e^2}{4\pi\epsilon_0 h} \frac{\gamma}{c} \quad (46)$$

From the above we can obtain the simple result that the number emitted per radian is:

$$N \frac{\rho}{c} = \frac{\gamma}{94.9} \quad (47)$$

which depends only on the energy,  $\gamma$ ; similarly the number emitted per metre is:

$$\frac{N}{c} = 6.2 B \quad (\text{electrons}) \quad (48)$$

which depends only on the magnetic field strength. It can be appreciated that the number of photons emitted per electron is a small quantity, and will therefore be subject to large statistical fluctuations. This has an important consequence for the beam dynamics that will be treated in the Chapter on "Quantum Excitation".

For a beam of electrons, the total number of photons emitted is given by:

$$\frac{dF}{d\theta} = \frac{N\rho}{c} \frac{I_b}{e} = 1.3 \cdot 10^{17} E [\text{GeV}] I_b [\text{A}] \quad \text{photons/sec/mrad horizontal} \quad (49)$$

In most cases therefore with average beam currents in the mA to several 100 mA range the total photon flux is large, with important consequences for the vacuum system (see section 5.2).

Two other expressions that will be of use in later Chapters are the mean and mean-square photon energy. The mean energy is:

$$\langle u \rangle = \frac{\int u n(u) du}{N} = \frac{P}{N} = \frac{8}{15\sqrt{3}} u_c \quad (50)$$

which is related by a numerical factor to the critical energy. Similarly the mean-square energy is related to the square of the critical energy:

$$\langle u^2 \rangle = \frac{\int u^2 n(u) du}{N} = \frac{8}{15\sqrt{3}} u_c^2 \int_0^\infty (u/u_c)^2 S(u/u_c) d(u/u_c) = \frac{11}{27} u_c^2 \quad (51)$$

## 5 . SYNCHROTRON RADIATION ASPECTS IN ELECTRON ACCELERATOR DESIGN

### 5.1 R.f. requirements

To make up for the losses due to the emission of SR the r.f. system must provide a sufficient accelerating voltage and sufficient power. There is an additional voltage requirement above that of the energy loss per turn in order to maintain sufficient beam lifetime, as will be described in a later Chapter on "Quantum Excitation". The total power needed from the r.f. system must also take into account the significant losses in the walls of the accelerating cavity.

The very strong dependence of the energy loss with energy is a limiting factor in the construction of increasingly high energy circular electron accelerators. Some reduction in the r.f. requirement can be achieved by reducing the bending magnet field strength, but at the expense of a larger and therefore more costly machine. It can be seen from Fig. 11 that in electron storage rings built so far, parameter optimization has resulted in a bending radius that increases roughly as  $E^2$  i.e. the field decreases as  $1/E$ . Thus the energy loss per turn,  $U_0$ , still increases as  $E^2$ . In LEP a field as low as 0.059 T is used at 55 GeV. Even so, 128 five-cell r.f. cavities are needed with sixteen 1 MW klystrons to power them [28]. To reach 100 GeV superconducting cavities will be used to reduce the power dissipated in the cavity. Because of the limitation imposed by SR emission attention is focussed on linear colliders for the next generation of high energy electron accelerators rather than circular machines.

As can be seen from Fig. 11, the situation is entirely different for proton machines. In this case SR emission is not a limiting factor and bending fields have increased by use of superconducting magnets.

### 5.2 Vacuum system

The power in the SR beams can be very high, up to several kW per metre of orbit length, and so water-cooled absorbers must be provided. The main gas load in the system is also caused by SR [29]: photons hitting the vacuum chamber produce photoelectrons which in turn desorb gas molecules from the surface. If the surface is clean before installation the main gas

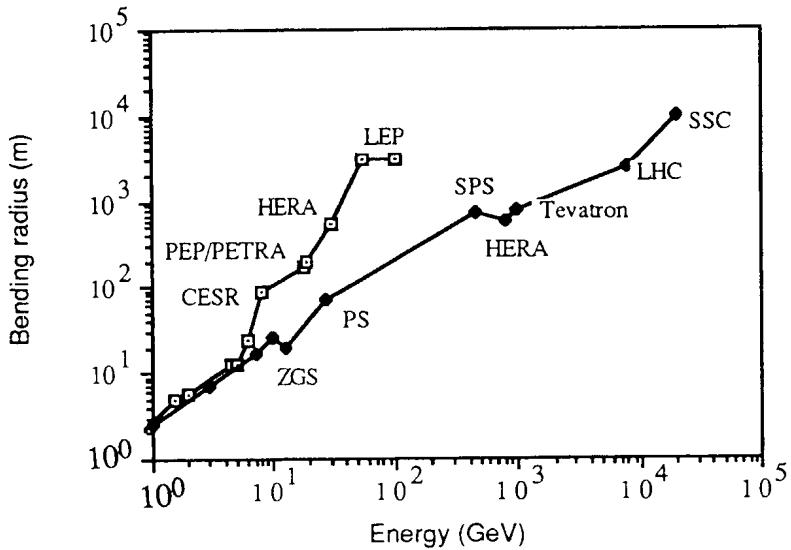


Fig. 11 Relationship between energy and bending radius for various circular electron and proton accelerators

molecules desorbed are H<sub>2</sub>, CO, CO<sub>2</sub> and CH<sub>4</sub>. The consequence of this is that the gas pressure will increase, and hence the beam lifetime will decrease, when the beam current increases. However, the desorption decreases as a function of the total beam 'dose', as the SR cleans the vacuum chamber surface. Typically it is assumed in the design of the pumping system that a dose of about 50 Amp-hours is required before nominal pressure and lifetime is reached.

### 5.3 Radiation shielding

When the critical energy of the photons becomes sufficiently large, scattered high energy photons can escape from the standard type of aluminium or stainless steel vacuum chamber and cause radiation damage to ring components (e.g. magnet coils and water hoses), failures of electronic components, production of ozone and nitric oxide in the air, and in the presence of humidity corrosion of the vacuum chamber by nitric acid. Problems of this type were experienced first in PETRA [30] and TRISTAN [31] when energies were raised above 15 GeV, which required lead shielding to be added around the chambers. Later higher-energy machines took this into account at the start. In the HERA electron ring a copper-bronze alloy was used for the chamber, and use was made of the dipole magnet yokes as extra shielding [32]. The LEP vacuum chamber is surrounded by a lead shield between 3 and 8 mm thick [28].

### 5.4 Electron-beam diagnostics

SR is widely used to form visible images of the electron beam [33] both for direct observation and qualitative measurement of beam size, position, stability and also bunch length. A typical arrangement is to reflect the visible part of the SR emitted by a bending magnet off a water-cooled mirror, through the radiation shield wall and into a diagnostics area. By focussing the light with a lens onto a screen a direct image of the electron beam can be formed. Linear photodiode arrays or CCD matrices can be used to obtain accurate information on the beam profile and a position sensitive detector can provide information about the positional stability of the beam. Bunch length information can be obtained from streak cameras or fast photomultipliers [34].

In storage rings used as synchrotron radiation sources, the stability of the position and angle of the radiation source points is very important. For this reason monitors which determine the position of the SR in the beam-lines have been developed. Using signals from these detectors feedback systems can then be used to stabilize the electron beam orbit.

## 6. SYNCHROTRON RADIATION SOURCES

As many as 28 electron storage rings are currently used for research using SR, and of these 17 are fully dedicated to SR. In addition, 16 new dedicated rings are under construction or commissioning, and several more are under study [35]. Not included in this list are about 10 smaller rings built exclusively as sources for X-ray lithography.

The properties of SR that make it so attractive as a research tool are as follows:

- high intensity or photon flux;
- continuous spectrum covering a broad range from the far infra-red to hard X-rays;
- small vertical angular divergence;
- small source size, determined mainly by the electron beam dimensions;
- high "brightness" and hence high partial coherence, resulting from the combination of small source size and divergence;
- polarization - linear in the orbit plane, with a circular component above and below the orbit plane;
- pulsed time structure, determined by that of the electron beam;
- calculable spectral intensity, allowing use as a calibrated source.

The wide range of applications cover the whole range of basic and applied science, and also include industrial (e.g. lithography and micro-mechanics) and medical (e.g. angiography) uses.

By convention the development of synchrotron radiation sources is usually described in terms of various "generations". The *first generation* includes synchrotrons and storage rings built initially for High Energy Physics and used "parasitically" for SR. The first experimental facilities for using SR were set up on the SURF and INS-SOR rings in 1963, followed by Frascati, DESY and Tantalus. Others followed, including the SPEAR ring (Stanford, USA), now dedicated to SR. The *second generation* rings are storage rings designed and built from the outset as dedicated SR sources. The first was the 300 MeV SOR-Ring in 1974, the first X-ray ring being the 2 GeV SRS (Daresbury, England). The *third generation* are newer dedicated rings with lower beam emittance (i.e. smaller beam sizes and divergences), and with many long straight sections into which special "insertion devices" (see below) can be placed. Third generation rings fall into three categories, according to the spectral region that they have been optimized for. Apart from one operating 0.8 GeV UV/VUV ring (SUPERACO, Orsay, France) the rest are either 1.2-2.0 GeV VUV/Soft X-ray rings or 6.0-8.0 GeV Soft/Hard X-rays rings.

The insertion devices that will form the main radiation sources in the third generation rings are magnetic devices with a field polarity that alternates along the electron beam trajectory [36]. A periodic deflection of the electron beam is produced, resulting in the emission of radiation with special properties that depend on the form of the magnetic field distribution. Compared to conventional bending magnet sources of SR, insertion devices can produce:

- higher photon energies;
- increased flux;
- increased brightness;
- different polarization characteristics.

Higher photon energies can be produced if the insertion device field, and hence critical photon energy, is larger than that of the bending magnets. If the device has a single high field pole, then it is often termed a *wavelength shifter*. The first such device was tested in the 240 MeV TANTALUS ring [37]. Several superconducting wavelength shifters are now operating with peak fields of 5-6 T. If the device has several magnet poles, then the output flux and brightness are increased accordingly and the device is generally called a *multipole wiggler*. If the parameters are such that interference effects are important, then very high brightness can be produced with a line, rather than continuous, spectrum; in this case the device is usually termed

an *undulator*. In the standard case where the insertion device deflects the particle in the horizontal plane the radiation is linearly polarized. Different polarization characteristics can be produced with more complicated helical, elliptical or non-sinusoidal electron trajectories [38].

Some aspects related to the design of third generation radiation sources and the effects of insertion devices on electron beam properties are considered in the Chapter on Quantum Excitation. Further details may be found in the Proceedings of a special CERN Accelerator School on Synchrotron Radiation Sources and Free Electron Lasers, Ref. [26].

## 7. SYNCHROTRON RADIATION FROM PROTONS

### 7.1 High energy proton rings

The equations governing the emission of Synchrotron Radiation are the same for both protons and electrons, however the total power radiated is inversely proportional to the 4th power of the rest mass, while the critical frequency of the spectrum varies as the inverse of the mass to the 3rd power. Thus, for the same energy and magnetic field electrons radiate  $10^{13}$  more power and with a critical frequency 6  $10^9$  times higher. Until recently the emission of SR in proton rings has not been of major concern, however, with the advent of the LHC and SSC this situation has changed. Table 2 lists the relevant parameters for some existing and future high energy proton machines.

**Table 2**  
Main parameters for various high energy proton accelerators

Ring	$E$ [TeV]	$\rho$ [m]	$B$ [T]	$U_0$ [keV]	$I_b$ [mA]	$P$ [kW]	$\epsilon_c$ [eV]
HERA	0.8	584	4.7	0.006	159	0.0009	0.003
Tevatron	1.0	754	4.4	0.011	2.5	0.0003	0.005
LHC	7.7	2568	10.0	10.7	851	9.1	63.7
SSC	20.0	10187	6.6	126	72	8.8	288

It can be seen that there are significant differences in the properties of the SR between the present and future generation of high energy proton machines. The consequences are particularly significant in the case of LHC and SSC since both rings employ superconducting magnets with a cold-bore vacuum tube [39, 40]. The SR power, although modest by electron ring standards, must be absorbed at cryogenic temperatures and represents a significant fraction of the heat input to the cryogenic system. To overcome this problem screens will be employed at a higher temperature than the vacuum pipe, on which the SR power will be absorbed more efficiently. A second problem is the effect on the vacuum pressure. The critical energy of the radiation in the LHC and SSC is sufficiently high to cause photo-desorption of gas molecules. These molecules will then be physisorbed by the screen which acts as a powerful cryopump. After a time a monolayer of adsorbed material can build up, whereupon the thermal vapour pressure of  $H_2$  will increase rapidly leading to a catastrophic pressure rise. Solutions to this problem, involving slots in the shield to pump the desorped gas molecules onto the colder vacuum tube surface, away from the impinging synchrotron radiation, are presently receiving detailed attention [41].

### 7.2 Modified synchrotron radiation properties

Although the spectrum of radiation emitted by a proton is in general described by the same equations as for an electron, an important difference can arise under certain conditions. In the derivation of section 3 it was assumed implicitly that the magnetic field was constant over the arc length seen by the observer (Fig. 2) i.e. over a distance  $L_o$  given by:

$$L_o \simeq \frac{\rho}{\gamma} = \frac{mc}{eB} \quad (52)$$

With a field of 1 T for example this corresponds to a length of 1.7 mm for electrons and 3.1 m for protons. Thus in general magnet lengths exceed the required value for the previous calculation to be valid. It has been shown however that an effect can be produced on the spectrum if at the edge of the magnet the field rises from zero to  $B$ , or falls from  $B$  to zero, in a distance  $\Delta L$  that is smaller than  $L_o$  [42]. In the case of electrons this condition is not normally met, but in the case of protons a typical magnet edge ( $\Delta L \sim 0.1$  m) is usually much smaller than the distance  $L_o$ . Since the effect is to shorten the emitted radiation pulse the resulting spectrum extends to higher critical frequency than that corresponding to the central magnet field,  $\omega_c$ , and is given approximately by:

$$\omega'_c \simeq \frac{L_o}{\Delta L} \omega_c \quad (53)$$

This "edge-effect" was observed for the first time on the SPS [43]. At 270 GeV and with a 1.2 T dipole field the standard critical wavelength of 16  $\mu\text{m}$  lies in the infra-red region of the spectrum. The magnet edge,  $\Delta L \sim 0.1$  cm, was much shorter than the value of  $L_o = 2.5$  m, and so the radiation emitted by the edge of the magnet had a critical wavelength of about 0.6  $\mu\text{m}$ , in the visible range. Later a special undulator magnet, which can be considered to be a series of short magnets with length smaller than  $L_o$ , was installed to produce visible radiation from both protons and anti-protons [44]. More recently a visible light beam monitor was also tested in the Tevatron using the edge-effect [45]. Critical frequencies in the LHC and SSC are already sufficiently high that visible radiation will be produced even without the edge-effect.

## ACKNOWLEDGEMENT

C.J. Bocchetta is thanked for many helpful suggestions and useful comments during the preparation of this and the following two chapters.

\* \* \*

## REFERENCES

- [1] A. Liénard, L'Éclairage Électrique 16 (1898) 5.
- [2] G.A. Schott, Electromagnetic Radiation, Cambridge University Press, 1912.
- [3] D. Iwanenko and I. Pomeranchuk, Phys. Rev. 65 (1944) 343.
- [4] J.P. Blewett, Phys. Rev. 69 (1946) 87.
- [5] J. Schwinger, Phys. Rev. 70 (1946) 798.
- [6] F.R. Elder et. al., Phys. Rev. 71 (1947) 829.
- [7] H.C. Pollock, Am. J. Phys. 51 (1983) 278.
- [8] J.P. Blewett, Nucl. Instr. Meth. Phys. Res. A266 (1988) 1.
- [9] F.R. Elder, R.V. Langmuir and H.C. Pollock, Phys. Rev. 74 (1948) 52.
- [10] J. Schwinger, Phys. Rev. 75 (1949) 1912.
- [11] D.A. Corson, Phys. Rev. 90 (1953) 748.

- [12] D.H. Tomboulian and P.L. Hartman, Phys. Rev. 102 (1956) 1423.
- [13] Iu.M. Ado and P.A. Cherenkov, Soviet Physics Doklady 1 (1956) 517.
- [14] N.H. Frank, Phys. Rev. 70 (1946) 177.
- [15] D. Bohm and L. Foldy, Phys. Rev. 70 (1946) 249.
- [16] J. Schwinger, Phys. Rev. 70 (1946) 798.
- [17] A.A. Sokolov and J.M. Ternov, J. Exp. Theor. Phys. USSR 25 (1953) 698.  
A.A. Sokolov and J.M. Ternov, Soviet Physics JETP 1 (1955) 227.
- [18] M. Sands, Phys. Rev. 97 (1955) 470.
- [19] A.A. Kolomenski and A.N. Lebedev, CERN Symposium (1956) p. 47.
- [20] I.G. Henry, Phys. Rev. 106 (1957) 1057.
- [21] F.A. Korolev et. al., Soviet Physics Doklady 6 (1960) 1011; Nuovo Cimento 18 (1960) 1033.
- [22] J.D. Jackson, Classical Electrodynamics, Wiley, New York (1962).
- [23] H. Bruck, Accélérateurs Circulaires de Particules, Presses Universitaires de France, Paris (1966).
- [24] A.A. Sokolov and I.M. Ternov, Synchrotron Radiation, Akademie-Verlag, Berlin (1968).
- [25] G.K. Green, Spectra and Optics of Synchrotron Radiation, Brookhaven National Laboratory, BNL 50522, April 1976.
- [26] A. Hofmann, Proc. CERN Accelerator School, Synchrotron Radiation and Free Electron Lasers, CERN 90-03, p. 115.
- [27] K.J. Kim, Characteristics of Synchrotron Radiation, AIP Conference Proceedings, 184 vol. 1, AIP, New York (1989) p. 565.
- [28] LEP Design Report Vol. II, CERN-LEP/84-01 (1984).
- [29] O. Grobner, Proc. CERN Accelerator School, General Accelerator Physics, CERN 85-19, p. 489.
- [30] The PETRA Storage Ring group, Proc. 11th Int. Conf. High Energy Accelerators, CERN, July 1980, Birkhäuser Verlag (1980) p. 16.
- [31] T. Momose et. al., Proc. 1st European Particle Accelerator Conference, Rome, June 1988, World Scientific (1989) p. 1284.
- [32] HERA-Proposal, DESY HERA 81-10, July 1981.
- [33] A. Hofmann and F. Meot, Nucl. Instr. Meth. 203 (1982) 483.
- [34] V.P. Suller, Proc. CERN Accelerator School, Synchrotron Radiation and Free Electron Lasers, CERN 90-03, p. 74.

- [35] V.P. Suller, Proc. 3rd European Particle Accelerator Conference, Berlin, March 1992, Editions Frontieres (1992) p. 77.
- [36] F. Ciocci, Proc. 3rd European Particle Accelerator Conference, Berlin, March 1992, Editions Frontieres (1992) p. 86.
- [37] W.S. Trzeciak, IEEE Trans. Nucl. Sci. NS-18 (1971) 213.
- [38] P. Elleaume, Proc. 1991 US Particle Accelerator Conference, IEEE 91CH3038-7, p. 1083.
- [39] W. Chou, Proc. 1991 US Particle Accelerator Conference, IEEE 91CH3038-7, p. 126.
- [40] J. Gomez-Goni et. al., Proc. 3rd European Particle Accelerator Conference, Berlin, March 1992, Editions Frontieres (1992) p. 1576.
- [41] A.G. Mathewson, these Proceedings.
- [42] R. Coisson, Opt. Comm. 22 (1977) 135.
- [43] R. Bossart et. al., Nucl. Instr. Meth. 164 (1979) 375.
- [44] J. Bosser et. al., J. Physique Lett. 45 (1984) L343.
- [45] A.A. Hahn and P. Hurh, Proc. 1991 U.S. Particle Accelerator Conference, IEEE 91CH3038-7, p. 1177.

## APPENDIX

### PRACTICAL SYNCHROTRON RADIATION FORMULAE

We first consider the most common case of electrons.

Equation (32) gives the energy radiated per unit solid angle per unit frequency interval during the passage of a single particle. In the case of a beam of particles the total energy per second, i.e. the radiated power (in Watts) is proportional to the number of particles that pass the observer per second:

$$\frac{d^2P}{d\Omega d\omega} = \frac{d^2W}{d\Omega d\omega} \frac{I_b}{e}$$

where  $I_b$  is the average beam current in Amps. The frequency interval can be expressed in any desired unit e.g. Hz, Ångstrom, wavenumbers etc. We use here the photon energy ( $\epsilon$ ), expressed in eV, which in practical units becomes (in Watts/mrad<sup>2</sup>/eV):

$$\frac{d^2P}{d\Omega d\epsilon} = 2.124 \cdot 10^{-3} E^2 [\text{GeV}] I_b [\text{A}] \left( \frac{\epsilon}{\epsilon_c} \right)^2 (1 + \gamma^2 \psi^2)^2 \left[ K_{2/3}^2(\xi) + \frac{\gamma^2 \psi^2}{1 + \gamma^2 \psi^2} K_{1/3}^2(\xi) \right]$$

where the critical photon energy is given by:

$$\epsilon_c [\text{keV}] = 0.665 E^2 [\text{GeV}] B[\text{T}] = 2.218 E^3 [\text{GeV}] / \rho$$

and the alternative critical wavelength as:

$$\lambda_c [\text{\AA}] = 18.64 / (E^2 [\text{GeV}] B[\text{T}]) = 5.589 \rho / E^3 [\text{GeV}]$$

It is also common to express SR intensities in terms of the number of photons per second, obtained by dividing the power in a given frequency interval by the appropriate photon energy  $\hbar\omega$ . Alternatively the power divided by  $\hbar$  gives the number of photons per second per unit *relative bandwidth*:

$$\frac{d^2F}{d\Omega d\omega/\omega} = \frac{d^2P}{d\Omega d\omega} \frac{1}{\hbar}$$

In practical units therefore (photons/s/0.1% bandwidth/mrad<sup>2</sup>):

$$\frac{d^2F}{d\Omega d\omega/\omega} = 1.325 \cdot 10^{13} E^2 [\text{GeV}] I_b [\text{A}] \left( \frac{\epsilon}{\epsilon_c} \right)^2 (1 + \gamma^2 \psi^2)^2 \left[ K_{2/3}^2(\xi) + \frac{\gamma^2 \psi^2}{1 + \gamma^2 \psi^2} K_{1/3}^2(\xi) \right]$$

and on-axis, ( $\psi = 0$ ):

$$\frac{d^2F}{d\Omega d\omega/\omega} = 1.325 \cdot 10^{13} E^2 [\text{GeV}] I_b [\text{A}] H_2(\epsilon / \epsilon_c)$$

Integrating over the vertical angle, the spectral distribution per unit horizontal angle becomes (in Watts/mrad/eV):

$$\frac{d^2P}{d\theta d\varepsilon} = 6.347 \cdot 10^{-3} E [\text{GeV}] I_b [\text{A}] S(\varepsilon / \varepsilon_c)$$

and expressed in terms of photon flux (photons/s/0.1% bandwidth/mrad horizontal):

$$\frac{d^2F}{d\theta d\omega / \omega} = 3.96 \cdot 10^{13} E [\text{GeV}] I_b [\text{A}] S(\varepsilon / \varepsilon_c)$$

The angular distribution of total power, Eq. (41), becomes (Watts/mrad<sup>2</sup>):

$$\frac{dP}{d\Omega} = 5.42 E^4 [\text{GeV}] B [\text{T}] I_b [\text{A}] \frac{1}{(1 + \gamma^2 \psi^2)^{5/2}} \left[ 1 + \frac{5\gamma^2 \psi^2}{7(1 + \gamma^2 \psi^2)} \right]$$

On-axis therefore the peak power density is given by:

$$\frac{dP}{d\Omega} = 5.42 E^4 [\text{GeV}] B [\text{T}] I_b [\text{A}] \quad \text{W/mrad}^2$$

Integrating over vertical angle gives the linear power density:

$$\frac{dP}{d\theta} = 4.22 E^3 [\text{GeV}] B [\text{T}] I_b [\text{A}] \quad \text{W/mrad horizontal}$$

The formulae above are valid for electrons. In the case of protons the results must be scaled by an appropriate factor  $\alpha^n$  where  $\alpha$  is the ratio of electron to proton mass i.e.  $\alpha = 1/1823$ . Thus,  $\varepsilon_c$  scales as  $\alpha^3$ ,  $\lambda_c$  as  $1/\alpha^3$ ,  $d^2P/d\Omega d\omega$  and  $d^2F/d\Omega(d\omega/\omega)$  as  $\alpha^2$ ,  $d^2P/d\theta d\omega$  and  $d^2F/d\theta(d\omega/\omega)$  as  $\alpha$ ,  $dP/d\Omega$  as  $\alpha^5$ ,  $dP/d\theta$  as  $\alpha^4$ .

# RADIATION DAMPING

*R.P. Walker*  
Sincrotrone Trieste, Italy

## Abstract

The basic formulae for the damping of the energy and betatron oscillations are derived. The results are applied to a number of examples of different lattice designs in which radiation damping effects are important. Methods of modifying and measuring the damping rates are also discussed.

## 1. INTRODUCTION

The loss of energy due to the emission of synchrotron radiation, and its replacement by the r.f. cavities, can give rise to a damping of the oscillations in energy and transverse displacement (synchrotron and betatron oscillations), a process known as "radiation damping". The only feature of the synchrotron radiation emission that is involved in this process is the rate of emission of energy, given in the previous Chapter on Synchrotron Radiation:

$$P = \frac{2}{3} \frac{e^2 c}{4\pi\epsilon_0} \frac{\beta^4 \gamma^4}{\rho^2} \quad (1)$$

From this can be obtained the total energy loss per turn,  $U_0$ , which in the case of an isomagnetic lattice (uniform bending radius in the bending magnets) is given by:

$$U_0 = \frac{e^2}{3\epsilon_0} \frac{\beta^3 \gamma^4}{\rho} \quad (2)$$

Because of the dependence on the fourth power of the rest mass the synchrotron radiation emission, and hence radiation damping effects, are only relevant for electrons at the energies of present-day accelerators. However, in the next generation of high energy proton accelerators the effects of radiation damping may start to become significant.

The process of radiation damping is important in many areas of electron accelerator operation:

- i) it can give rise to a stable (Gaussian) distribution of transverse and longitudinal beam dimensions due to an equilibrium between the competing forces of radiation damping and "quantum excitation" – the growth of oscillation amplitudes due to the discrete emission of radiation quanta;
- ii) it permits an efficient multi-cycle injection scheme to be employed in storage rings, by allowing the beam dimensions to damp in size between injection pulses;
- iii) it allows large beam dimensions, produced in a linac for example, to be reduced in specially designed "damping rings";
- iv) it helps to counteract beam growth due to various processes such as intra-beam scattering and collective instabilities.

In this chapter the basic formulae for the damping of the energy and betatron oscillations are derived, following closely the treatment in earlier texts [1–3]. The main results are illustrated by a number of examples of different lattice designs in which radiation damping effects are important. Methods of modifying the damping rates in a given ring are then discussed and finally techniques for the measurement of the damping rates are considered. The

following Chapter deals with the related quantum excitation process and the derivation of the equilibrium beam dimensions.

## 2 . ENERGY OSCILLATIONS

Figure 1 shows the accelerating voltage, and hence the energy gain, in an r.f. cavity as a function of the time of arrival of an electron. The particle which arrives on every turn at the correct time (and hence phase with respect to the r.f. voltage) in order to make up the loss due to synchrotron radiation ( $U_0$ ) is called the synchronous particle, and its energy is the nominal energy of the design orbit,  $E_0$ . An electron with a higher energy will in general travel on a longer path and therefore arrive later at the cavity. It can be seen from Fig. 1 that such a particle will receive less energy at the cavity, which therefore compensates for the energy deviation. Similarly, a lower energy particle travels on a shorter path, arrives earlier at the cavity and therefore has a higher energy gain. This describes the usual stable oscillations in energy and time that occur about the synchronous point, which are analysed in more detail in Ref. [4]. If in addition the energy loss due to synchrotron radiation increases with the energy of the particle, then it can be seen that this will provide a damping of the oscillations.

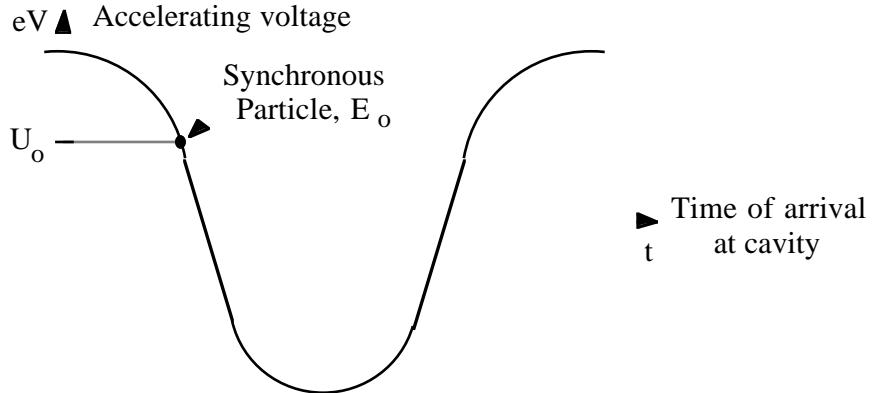


Fig. 1 Variation of accelerating voltage in an r.f. cavity as a function of electron arrival time

We now consider this damping process in more detail. The standard terminology will be used which refers to the time displacement of an electron with respect to the synchronous particle, or equivalently to the centre of the bunch, as shown in Fig. 2. In this description an electron which is ahead of the synchronous electron by a distance  $\Delta s$  has a positive time displacement  $\tau = \Delta s/c$ . An electron with a positive energy deviation  $\varepsilon = E - E_0$  has a larger orbit length ( $L$ ) and hence orbit period ( $T$ ) with respect to the synchronous particle (denoted by the subscript  $o$ ) given by:

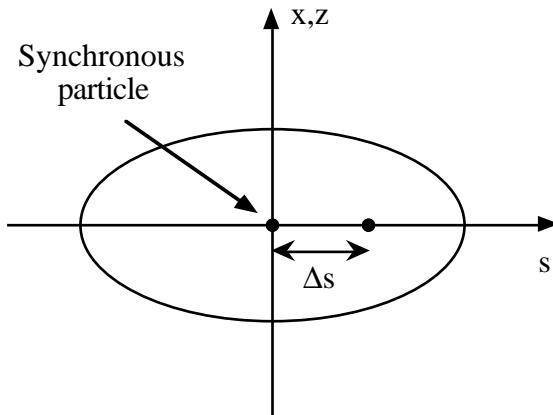


Fig. 2 Location of the synchronous particle and an electron with a positive time displacement in an electron bunch

$$\frac{\Delta T}{T_0} = \frac{\Delta L}{L_0} = \alpha \frac{\varepsilon}{E_0}$$

where  $\alpha$  is the momentum compaction factor, neglecting the relativistic factor which is negligible for electrons [4]. We assume that changes in energy and time displacement occur slowly with respect to the orbit period, which permits use of a differential notation:

$$\frac{d\tau}{dt} = -\alpha \frac{\varepsilon}{E_0} \quad (2)$$

Considering the energy equation, in one turn an electron loses an energy  $U(\varepsilon)$  and gains from the r.f. cavity  $eV(\tau)$ ; the net change is therefore:

$$\Delta E = eV(\tau) - U(\varepsilon)$$

and so on average:

$$\frac{d\varepsilon}{dt} = \frac{eV(\tau) - U(\varepsilon)}{T_0}$$

Taking the derivative of the above then gives:

$$\frac{d^2\varepsilon}{dt^2} = \frac{e}{T_0} \left( \frac{dV(\tau)}{d\tau} \frac{d\tau}{dt} \right) - \frac{1}{T_0} \left( \frac{dU(\varepsilon)}{d\varepsilon} \frac{d\varepsilon}{dt} \right)$$

Inserting Eq. (2) gives:

$$\frac{d^2\varepsilon}{dt^2} + \frac{1}{T_0} \frac{dU(\varepsilon)}{d\varepsilon} \frac{d\varepsilon}{dt} + \frac{e}{T_0} \frac{\alpha}{E_0} \frac{dV(\tau)}{d\tau} \varepsilon = 0 \quad (3)$$

For small oscillations it can be assumed that the accelerating voltage varies linearly with respect to the arrival time around that of synchronous particle:

$$eV(\tau) = U_0 + e\dot{V}_0 \tau$$

where  $\dot{V}_0$  is the slope of the accelerating voltage  $dV(\tau)/d\tau$  at  $\tau = 0$ . Using this expression Eq. (3) can be written as follows:

$$\frac{d^2\varepsilon}{dt^2} + 2\alpha_\varepsilon \frac{d\varepsilon}{dt} + \Omega^2 \varepsilon = 0$$

where:

$$\alpha_\varepsilon = \frac{1}{2} \frac{dU}{dT_0} \frac{dU}{d\varepsilon} \quad (4)$$

$$\Omega^2 = \frac{e}{T_0} \dot{V}_0 \frac{\alpha}{E_0}$$

This can be recognised as the usual equation of harmonic motion for the energy oscillations [4] with an additional damping term. Assuming that the damping rate  $\alpha_\varepsilon$  is small with respect to the oscillation frequency  $\Omega$ , the solution can be written as follows:

$$\varepsilon(t) = A e^{-\alpha_\varepsilon t} \cos(\Omega t - \phi)$$

$$\tau(t) = \frac{-\alpha}{E_0 \Omega} A e^{-\alpha \varepsilon t} \sin(\Omega t - \phi)$$

where  $A$  and  $\phi$  are constants determined by the initial conditions. It can be seen that as anticipated the damping rate depends on the change in energy loss with energy deviation ( $dU(\varepsilon)/d\varepsilon$ ).

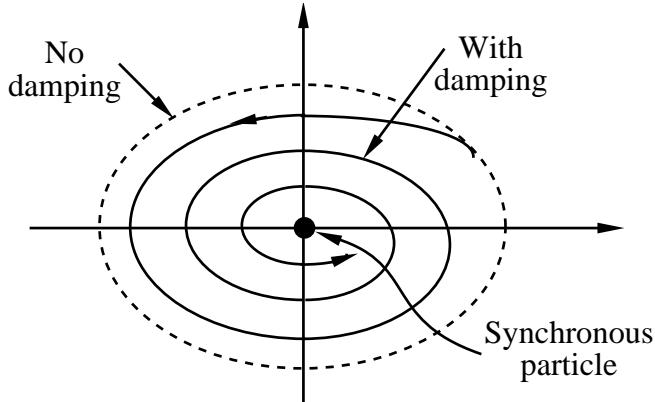


Fig. 3 Illustration of energy oscillations with and without radiation damping

Figure 3 illustrates the above solutions. In the absence of damping an electron executes a harmonic oscillation in energy and time with a fixed amplitude that is represented by an ellipse of a given size. With positive damping the particle spirals slowly towards the fixed point, namely the synchronous particle.

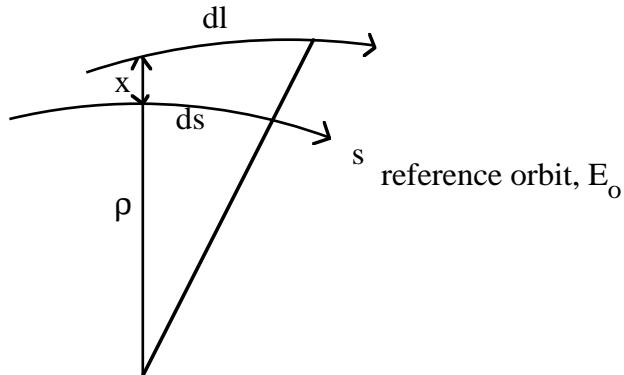


Fig. 4 Elements of the reference orbit and a displaced trajectory

We now consider how to calculate the damping rate ( $\alpha_\varepsilon$ ) from the rate of change of energy loss with energy ( $dU/d\varepsilon$ ). When the energy deviates from the nominal energy  $E_0$  the energy loss changes because of several factors. Firstly, the energy loss is itself a function of energy and secondly because the orbit deviates from the reference orbit there may be a change in magnetic field and a change in the path length. Figure 4 shows a curved element of the design orbit in the horizontal plane for a particle of the nominal energy  $E_0$ , where the radius of curvature is  $\rho$ . Also shown is the trajectory of another particle with transverse displacement,  $x$ . In general the path length for the elements are related as follows:

$$\frac{dl}{ds} = 1 + \frac{x(s)}{\rho} \quad (5)$$

For an off-energy particle the closed orbit is defined by:

$$x(s) = D(s) \frac{\varepsilon}{E_0} \quad (6)$$

where  $D(s)$  is the dispersion function. In this case therefore:

$$\frac{dl}{ds} = 1 + \frac{D}{\rho} \frac{\varepsilon}{E_0} \quad (7)$$

The energy radiated per turn is defined as the integral of the radiated power ( $P$ ) around the off-energy orbit:

$$U(\varepsilon) = \frac{1}{c} \oint P dl$$

Using Eq. (7) for the path length this can be expressed as an integral over  $s$ :

$$U(\varepsilon) = \frac{1}{c} \oint P \left( 1 + \frac{D}{\rho} \frac{\varepsilon}{E_0} \right) ds$$

We now expand  $P$  as a function of energy and transverse displacement, given the fact that  $P$  is proportional to  $E^2$  and  $B^2(x)$ , Eq. (1):

$$P(s) = P_0 + \frac{2P_0}{E_0} \varepsilon + \frac{2P_0}{B_0} \frac{dB}{dx} x \quad (8)$$

where  $P_0(s)$  is the power radiated on the design orbit, corresponding to the field  $B_0$ . Inserting in the expression for  $U(\varepsilon)$  together with Eq. (6) and keeping only linear terms in  $\varepsilon$  we obtain:

$$U(\varepsilon) = \frac{1}{c} \oint \left( P_0 + \frac{2P_0}{E_0} \varepsilon + \frac{dB}{dx} D \frac{\varepsilon}{E_0} + \frac{P_0 D}{\rho} \frac{\varepsilon}{E_0} \right) ds$$

For the damping rate we require the derivative:

$$\frac{dU(\varepsilon)}{d\varepsilon} = \frac{1}{c} \oint \left( \frac{2P_0}{E_0} - \frac{2P_0 k \rho D}{E_0} + \frac{P_0 D}{\rho E_0} \right) ds$$

where the usual focusing parameter  $k$  for gradient fields has been introduced ( $k\rho = [-dB/dx]/B_0$ ). Since the integral of  $P_0/c$  around the design orbit is  $U_0$ , we obtain:

$$\frac{dU(\varepsilon)}{d\varepsilon} = \frac{2U_0}{E_0} + \frac{1}{cE_0} \oint P_0 D (1/\rho - 2k\rho) ds$$

and hence the equation for the damping rate can be expressed in the following standard form:

$$\alpha_\varepsilon = \frac{1}{2T_0} \frac{dU}{d\varepsilon} = \frac{1}{2T_0} \frac{U_0}{E_0} (2+D)$$

where:

$$D = \frac{1}{cU_0} \oint P_0 D (1/\rho - 2k\rho) ds \quad (9)$$

Using the fact that  $P_0$  depends on  $1/\rho^2$ , the important parameter  $D$  can be expressed in the following standard forms, involving only integrals over various lattice functions:

$$D = \frac{\oint D / \rho (1/\rho^2 - 2k) ds}{\oint 1/\rho^2 ds} = \frac{\oint D(1-2n) / \rho^3 ds}{\oint 1/\rho^2 ds} \quad (10)$$

It is clear that  $D$  is a dimensionless number, with contributions only from the ring bending magnets ( $1/\rho \neq 0$ ). One term involves both bending and focusing fields ( $k/\rho \neq 0$ ) which is present in "combined function" or "synchrotron magnets". For these magnets it is convenient to define a field index,  $n$ :

$$n = -\frac{dB}{dx} \frac{\rho}{B_0} = k\rho^2$$

which appears in the second expression above for  $D$ .

From the expression for the damping rate, Eq. (4), we recall that  $dU/d\varepsilon$  must be positive for the oscillations to be damped and hence  $D > -2$ . In the most common case of a "separated function" lattice (as will be seen later)  $D$  is a small positive number, in which case  $dU/d\varepsilon$  is determined only by the  $E^2$  dependence of  $P$ . In this case we have the result that:

$$\tau_\varepsilon = \frac{1}{\alpha_\varepsilon} \approx \frac{T_0 E_0}{U_0} \quad (11)$$

i.e. the damping time is approximately the time it would take for an electron to radiate away all its energy (at constant rate), a useful and easily remembered result.

We conclude this section with a table giving various parameters connected with the energy oscillations for two widely different electron machines at CERN, the EPA [5] and LEP [6]. It can be seen that in both cases the damping time is much longer than the synchrotron oscillation and orbit periods, justifying the approximations used in the derivation above. Finally, we note that for protons even at the high energy expected at the SSC (20 TeV) the damping time is still extremely long, about 12 hours.

**Table 1**

Energy oscillation parameters for two electron storage rings

	EPA [5]	LEP [6]
Energy, $E_0$ (GeV)	0.6	55
Energy loss per turn, $U_0$ (keV)	8	$260 \cdot 10^3$
Orbit period, $T_0$ (μs)	0.42	89
Synchrotron oscillation period (ms)	0.27	1
Synchrotron oscillation damping time, $\tau_\varepsilon$ (ms)	64	18

### 3 . BETATRON OSCILLATIONS

We consider now the damping of the betatron oscillations, starting with the more simple case of the vertical plane.

#### 3.1 Vertical plane

It is convenient to use the following approximate form for the vertical betatron oscillations:

$$z = A \cos(\phi(s) + \phi_0) \quad z' = \frac{-A}{\beta} \sin(\phi(s) + \phi_0)$$

where  $A$  is the normalised amplitude of the oscillation:

$$A^2 = z^2 + (\beta z')^2 \quad (12)$$

However, it is easy to show that the same result is obtained if the complete form for the amplitude is used:

$$A^2 = \gamma z^2 + 2\alpha z z' + \beta z'^2$$

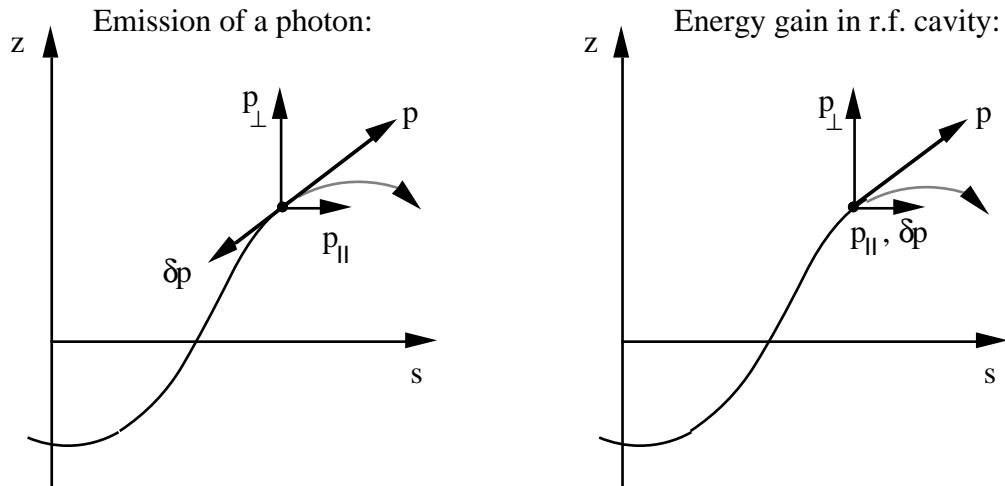


Fig. 5 Effect of energy loss and energy gain processes on the electron momentum

We wish to consider the effect on the oscillation amplitude  $A$  of energy loss due to synchrotron radiation and energy gain in the r.f. cavities. These processes are illustrated in Fig. 5, occurring at an arbitrary point with respect to the phase of the betatron oscillation. It can be seen that since photons are emitted in the direction of the motion of the electron, there is a change in the value of the momentum, but no change in angle  $z'$ . On the other hand, in the r.f. cavity there is an increase in the longitudinal component of the momentum ( $p_{||}$ ) which therefore reduces the angle. Since  $z' = p_{\perp}/p_{||}$ , after the cavity we have

$$z' + \delta z' = \frac{p_{\perp}}{p_{||} + \delta p} \approx z'(1 - \frac{\delta p}{p})$$

and hence:

$$\delta z' = -z' \frac{\delta \epsilon}{E_0}$$

Using Eq. (12) the change in oscillation amplitude is given by:

$$A \delta A = \beta^2 z' \delta z' = -\beta^2 z'^2 \frac{\delta \epsilon}{E_0}$$

Averaging over all possible phases of the oscillation at the time the electron passes through the cavity,  $\langle z'^2 \rangle = A^2/2\beta$ , we have:

$$\frac{\langle \delta A \rangle}{A} = -\frac{1}{2} \frac{\delta \epsilon}{E_0}$$

Since the gain in energy over one turn is small compared to the electron energy we can average over one turn to obtain:

$$\frac{\Delta A}{A} = \frac{-U_0}{2 E_0} \quad (13)$$

The motion is therefore exponentially damped ( $\exp -\alpha_z t$ ) with a time constant  $\alpha_z$  given as follows:

$$\alpha_z = -\frac{1}{A} \frac{dA}{dt} = \frac{U_0}{2 E_0 T_0}$$

which is one half of the approximate value for the energy oscillations derived in the previous section, Eq. (11).

### 3.2 Horizontal plane

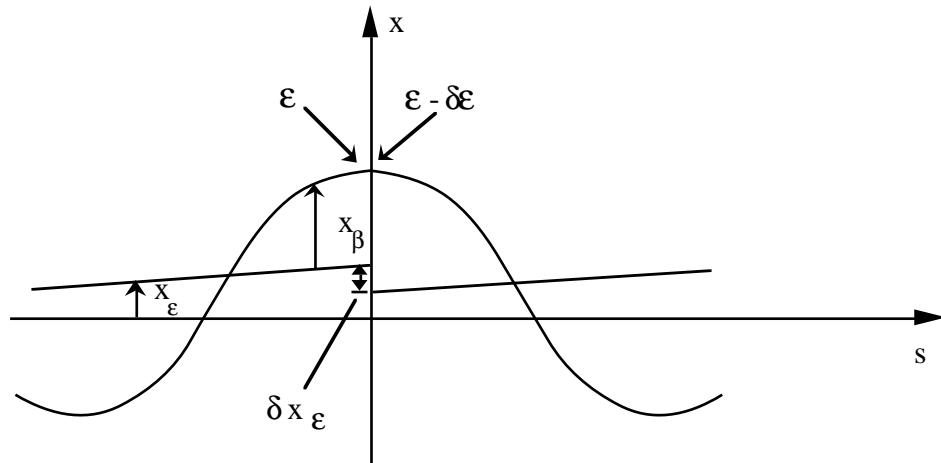


Fig. 6 Effect of energy loss on the off-energy orbit and betatron motion in the horizontal plane

The same process as above occurs in the horizontal plane also, but there is an additional effect due to the emission of synchrotron radiation at points where there is finite dispersion (which is usually zero in the vertical plane). As before, there is no change in  $x$  or  $x'$  due to the radiation emission (see Fig. 6), however the change in energy implies a change in the off-energy orbit ( $x_\epsilon = D(s) \epsilon/E_0$ ) and hence an equal and opposite change in the betatron amplitude  $x_\beta$ , since  $x = x_\epsilon + x_\beta$ . We have therefore:

$$\delta x_\beta = -\delta x_\epsilon = -D \frac{\delta \epsilon}{E_0}$$

Similarly, the change in angle of the betatron oscillation is given by:

$$\delta x_{\beta'} = -\delta x_{\epsilon'} = -D' \frac{\delta \epsilon}{E_0}$$

The change in the oscillation amplitude  $A$  where:

$$A^2 = x_\beta^2 + (\beta x'_\beta)^2$$

is therefore given as follows:

$$A \delta A = x_\beta \delta x_\beta + \beta^2 x'_\beta \delta x'_\beta = - (D x_\beta + \beta^2 D' x'_\beta) \frac{\delta \epsilon}{E_0} \quad (14)$$

If the rate of energy loss were constant then averaging over the betatron phase would yield no net increase in amplitude, however, this is not the case and it is necessary to include the variation of energy loss with  $x_\beta$ . Since the energy loss in a small element is given by  $\delta \epsilon = -P/c \delta l$ , where (Eq. (5)):

$$\frac{\delta l}{\delta s} = 1 + \frac{x_\beta}{\rho}$$

and expressing  $P$  as a function of  $x_\beta$ , as in Eq. (8):

$$P = P_0 + \frac{2P_0}{B_0} \frac{dB}{dx} x_\beta = P_0 (1 - 2k\rho x_\beta)$$

gives the result:

$$\delta \epsilon = -\frac{P_0}{c} \left( 1 - 2k\rho x_\beta + \frac{x_\beta}{\rho} \right) \delta s$$

Combining with Eq. (14) and averaging over the betatron phase, given that  $\langle x_\beta \rangle = 0$ ,  $\langle x'_\beta \rangle = 0$ ,  $\langle x_\beta x'_\beta \rangle = 0$ , and  $\langle x_\beta^2 \rangle = A^2/2$  gives:

$$\frac{\langle \delta A \rangle}{A} = \frac{P_0}{2cE_0} D (1/\rho - 2k\rho) \delta s$$

Over one turn therefore:

$$\frac{\Delta A}{A} = \frac{1}{2cE_0} \oint P_0 D (1/\rho - 2k\rho) ds = \frac{U_0 D}{E_0} \frac{1}{2}$$

using the earlier definition of the quantity  $D$ , Eq. (9).

In general, since  $D$  is usually positive, this would give rise to an increase in oscillation amplitude, however when the effect of the damping that occurs due to the energy gain in the r.f. cavities is added, as in the vertical plane Eq. (13), we have:

$$\frac{\Delta A}{A} = \frac{-U_0}{2E_0} (1-D)$$

and hence the damping rate is as follows:

$$\alpha_x = -\frac{1}{A} \frac{dA}{dt} = \frac{U_0}{2E_0 T_0} (1-D) \quad (15)$$

### 3.3 Origin of the damping of the betatron motion

It is interesting to note that the main damping effect of the betatron motion described in section 3.1 appears to occur due to the energy gain in the r.f. cavities, not due to the energy loss, and as a result it has been remarked that the term "radiation damping" is somewhat

inappropriate. However, it can be seen from Fig. 5 that in fact the opposite is true if the canonically conjugate variables of position and momentum are used rather than the more usual position and angle, since the transverse momentum,  $p_\perp$ , is reduced when a photon is emitted but is unchanged in the r.f. cavity [7]. The choice of variables therefore determines the apparent location of the damping effect, however the final result is the same.

#### 4. DAMPING PARTITION AND THE ROBINSON THEOREM

The results obtained in the previous two sections may be summarized as follows:

$$\alpha_i = \frac{J_i U_0}{2 E_0 T_0} \quad (16)$$

where  $i$  represents  $x$ ,  $z$  or  $\epsilon$  and  $J_i$  are the Damping Partition Numbers:

$$J_x = 1-D \quad J_z = 1 \quad J_\epsilon = 2+D$$

so called because the sum of the damping rates for the three planes is a constant:

$$J_x + J_z + J_\epsilon = 4 \quad (17)$$

a result known as the Robinson Theorem [8]. For damping in all planes simultaneously it is required that all  $J_i > 0$  and hence that  $-2 < D < 1$ .

We have obtained the above result for the total damping explicitly by analysing each oscillation mode independently, however it may be obtained in a more general and direct way using the following method [8]. The general transverse and longitudinal motion of a particle with respect to that of the synchronous particle on the design orbit may be described using 6x6 transfer matrices, relating particle coordinates at some initial position  $s_1$  to those at some later position  $s_2$  as follows:

$$\begin{bmatrix} x \\ x' \\ z \\ z' \\ \epsilon/E_0 \\ \tau \end{bmatrix}_{s_2} = \mathbf{M}(s_2, s_1) \begin{bmatrix} x \\ x' \\ z \\ z' \\ \epsilon/E_0 \\ \tau \end{bmatrix}_{s_1}$$

Since the elements of  $\mathbf{M}$  are real then the eigenvalues of the one-turn matrix  $\mathbf{M}(s+L, s)$  can be written as three complex conjugate pairs  $\exp(-\alpha'_j \pm i\beta'_j)$  with  $j = 1, 2, 3$ . Using the fact that the determinant of a matrix is the product of its eigenvalues [9] we have:

$$\det \mathbf{M}(s+L, s) = \exp \left( -\sum_{j=1}^3 2\alpha'_j \right) \cong 1 - \sum_{j=1}^3 2\alpha'_j \quad (18)$$

since  $\alpha'_j \ll 1$ . The amplitudes of the three oscillation modes vary as  $\exp(-\alpha_j t)$  where  $\alpha_j = \alpha'_j/T_0$  i.e.  $\alpha_j$  are the damping rates.

Considering a general infinitesimal element of orbit between  $s$  and  $s+ds$ , the matrix can be written:

$$\mathbf{M}(s+ds, s) = \mathbf{I} + \delta\mathbf{M}$$

where  $\mathbf{I}$  is the identity matrix. Since all elements of  $\delta\mathbf{M}$  are small, it can be shown that:

$$\det \mathbf{M}(s+ds, s) \approx 1 + T_r(\delta\mathbf{M})$$

where  $T_r$  represents the trace of the matrix. In the absence of energy loss and gain the determinant of  $\mathbf{M}$  is equal to unity. The only diagonal terms in  $\delta\mathbf{M}$  therefore are those calculated earlier representing changes in  $x'$  and  $z'$  due to gain of energy  $\delta\varepsilon_1$ :

$$\delta x' = -\frac{\delta\varepsilon_1}{E_0} x' , \quad \delta z' = -\frac{\delta\varepsilon_1}{E_0} z'$$

as well as that for  $\varepsilon/E_0$  due to energy loss:

$$\delta\varepsilon = -2\frac{\delta\varepsilon_2}{E_0} \varepsilon$$

since the rate of emission is proportional to  $E^2$ . We have therefore:

$$\det \mathbf{M}(s+ds, s) = 1 - 2\frac{\delta\varepsilon_1}{E_0} - 2\frac{\delta\varepsilon_2}{E_0}$$

For the one-turn matrix, since the determinant of a product of matrices is the product of the determinants and since the total energy gain and the total energy loss are equal to  $U_0$ , we have that:

$$\det \mathbf{M}(s+L, s) = 1 - \frac{4 U_0}{E_0}$$

irrespective of the location of the energy loss and gain. Combining with Eq. (18) gives the final result:

$$\sum_{j=1}^3 \alpha_j = \frac{2}{E_0} \frac{U_0}{T_0}$$

which is identical to the one obtained earlier, Eqs. (16) and (17). The present derivation however shows that the result is independent of the nature of the magnetic and electric field distributions acting on an electron, provided that they are determined *a priori*, i.e no beam induced fields are included. It is valid therefore even in the case of linear coupling between the horizontal and vertical planes, and when there is bending in the vertical plane. In the absence of these factors the matrices for the  $(z, z')$  and  $(x, x', \varepsilon/E_0, \tau)$  motion and may be treated separately, giving the result:

$$\alpha_x + \alpha_\varepsilon = \frac{3}{2} \frac{U_0}{E_0} \frac{T_0}{T_0}$$

or equivalently,

$$J_x + J_\varepsilon = 3.$$

## 5 . RADIATION DAMPING ASPECTS IN VARIOUS LATTICE DESIGNS

### 5.1 Weak focusing lattices

Early accelerators employed "weak focusing" magnets that provided focusing in both planes simultaneously for which the field index must lie in the range  $0 < n < 1$  [10]. There is a

further constraint on the field index in order that the motion is damped in all three planes. To derive this we first write the expression for  $D$  in a form that is valid in the case of an isomagnetic lattice:

$$D = \frac{1}{2\pi\rho} \oint (1 - 2n) \frac{D}{\rho} ds \quad (19)$$

We leave  $\rho$  inside the integral to indicate that it includes only the bending magnets and not any straight sections. We can simplify this by making use of the expression that defines the dispersion function:

$$D'' = \left( k - \frac{I}{\rho^2} \right) D + \frac{I}{\rho}$$

from which it follows by integration that:

$$\oint \left( 1/\rho^2 - k \right) D ds = \oint (1/\rho) ds$$

If the focusing is due entirely to combined function magnets, with field index  $n = kp^2$ , then the above may be written in the isomagnetic case as follows:

$$\oint (1-n) \frac{D}{\rho} ds = 2\pi\rho \quad (20)$$

If we now include the fact that the field index is also constant in the bending magnets, then combining Eqs. (19) and (20) gives:

$$D = \frac{1-2n}{1-n}$$

It follows that the damping partition numbers are then given by:

$$J_x = \frac{n}{1-n} \quad J_z = 1 \quad J_\epsilon = \frac{3-4n}{1-n}$$

and so for damping in all three planes  $0 < n < 0.75$ . The fact that energy oscillations become undamped for  $n > 0.75$  was appreciated even before the first observation of synchrotron radiation [11-13].

A present day example of this type of lattice is the NBS 250 MeV storage ring which is used as a synchrotron radiation facility (SURF). Originally however the ring was operated as a 180 MeV synchrotron with a field index of 0.8; when it was converted for use as a storage ring extra gradient coils were added to lower the field index to 0.6 in order to obtain the necessary damping of all oscillation modes [14]. A more recent example is the compact superconducting synchrotron radiation source AURORA whose field index varies in the range 0.3–0.7 as the energy is varied between 150 and 650 MeV [15].

## 5.2 Strong focusing, combined function

Several early types of alternating gradient or "strong focusing", synchrotrons were constructed using magnets with combined bending and focusing fields, for example the CEA and DESY I electron synchrotrons, as well as the PS proton synchrotron. If there are no separate focusing fields ( $k \neq 0$  only if  $I/\rho \neq 0$ ) then combining Eqs. (19) and (20) above, gives in the isomagnetic case:

$$D = 2 - \frac{\oint (D/\rho) ds}{2\pi\rho} = 2 - \frac{\alpha L}{2\pi\rho}$$

where  $\alpha$  is the momentum compaction factor. Since  $\alpha$  is usually small it may be seen that  $D \sim 2$  and hence  $J_x \sim -1$ ,  $J_z = 1$  and  $J_\varepsilon \sim 4$ . In the case of a combined function lattice therefore the betatron motion is anti-damped in the horizontal plane [16–18]. Electron synchrotrons can however be built with a combined function lattice, provided the growth that occurs in the horizontal beam size is acceptable.

In order to overcome the anti-damping of the combined function lattice various correction methods have been proposed [8,16,17,19,20], some of which are discussed in Section 6.

### 5.3 Strong focusing, separated function

It has been shown that radial damping can be achieved in a combined function lattice by using focusing and defocusing magnets of slightly different strength [8, 21]. The most common lattice arrangement however which produces damping in all three planes, is the so-called separated function lattice i.e. one in which the functions of bending and focusing are divided in separate dipole and quadrupole magnets [22,23]. One possibility may be seen directly from Eq. (10). It is clear that with a value of  $n = 0.5$  in the bending magnets  $D = 0$ , and hence  $J_x = J_z = 1$ ,  $J_\varepsilon = 2$ , irrespective of additional quadrupole magnets which may be arranged to produce an alternating gradient structure. Such an approach was taken in the design of both the ACO and ADONE storage rings.

In the case of zero field gradient in the dipole magnets, it may be seen from Eq. (10) that in the isomagnetic case we have:

$$D = \frac{\oint (D/\rho) ds}{2\pi\rho} = 2 - \frac{\alpha L}{2\pi\rho}$$

The value of  $D$  in this case results only from the path length effect in the dipole magnets, which is usually very small. In all of the above analysis we have assumed that the bending magnets have a sector geometry, however, only small modifications usually result in the case of non-zero entrance and exit angles. In the special case of a lattice with parallel edged dipole magnets it may be shown that the effective field gradient at the entrance and exit of the magnet cancels the path length effect exactly, resulting in  $D = 0$ .

In separated function lattices therefore  $J_x \sim 1$ ,  $J_z = 1$  and  $J_\varepsilon \sim 2$ , and so the motion is damped in all three planes. This type of lattice is now generally used not only for storage rings, but also for synchrotrons since this also leads to smaller beam sizes. The difference between the two lattice types may be illustrated by the performance of the original DESY I synchrotron (combined function) and the later DESY II (separated function) [24], shown in Fig. 7. In DESY I the increase in horizontal beam size after an initial period of adiabatic damping is due to fact that the horizontal motion is anti-damped. A high repetition rate of 50 Hz was necessary in this case in order to limit the growth of the beam size. On the other hand, in DESY II the beam size approaches the equilibrium value even for widely different injected beam sizes and a much slower repetition rate could be used (12.5 Hz).

### 5.4 Damping time and injection energy

A common type of injection scheme for electron storage rings is multi-cycle injection, in which the injected beam damps in size due to radiation damping in the interval between injections so preventing loss on the injection septum magnet. In this way a high current can be

accumulated without needing a very high performance injector. The maximum possible injection rate depends to some extent on the damping time for the plane in which the injection is performed, usually the horizontal. This is particularly important when a ring is being filled at a lower energy than its final operating value since the damping time varies rapidly with energy,  $\sim 1/E^3$ .

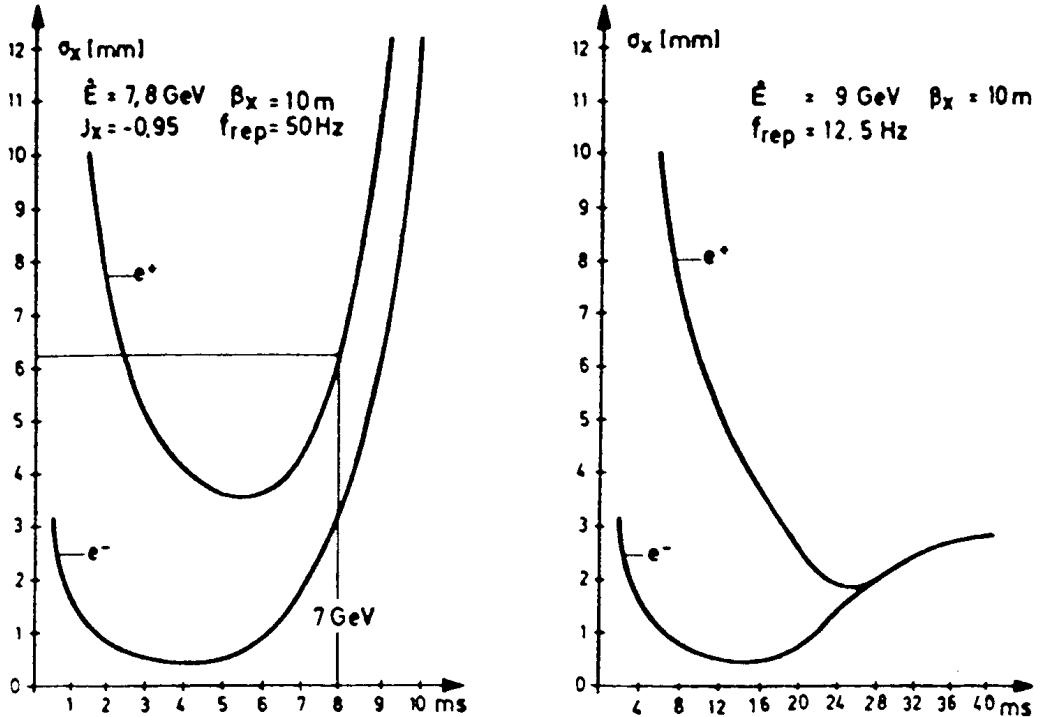


Fig. 7 Variation of horizontal beam size with time during the acceleration cycle in the DESY I (left) and DESY II (right) synchrotrons [24]

The importance of injection energy is illustrated by the unique system of beam storage that was employed at the CEA when it was operated as an electron storage ring with a low injection energy of 260 MeV [25]. In order to increase the current that could be accumulated the energy was cycled repeatedly between injection energy and 2.1 GeV, so that sufficient radiation damping could occur at the higher energy between successive injections.

The topic of injection energy is particularly relevant in the field of modern compact sources of synchrotron radiation [26]. Since the critical wavelength of the radiation at the operating energy varies as  $\rho/E_0^3$ , it follows that the same value can be obtained with a lower operating energy using superconducting magnets with a smaller bending radius than conventional magnets. This tends to reduce the overall circumference and so make the ring more compact. In addition, since the damping time at the injection energy varies as  $T_0\rho/E_i^3$  it follows that a lower injection energy may be used while maintaining the same damping time, which permits a more compact and cheaper injector to be used.

**Table 2**

Damping times and injection rates in some electron storage rings with low energy multi-cycle injection

Injection energy (MeV)	COSY [29]	MAX [27]	ALADDIN [30]
Damping time (s)	50	100	100
	2.5	2.5	13.6

Injection rate (Hz)	10	10	1.25
---------------------	----	----	------

Many other factors, however, affect the injection process at low energy, such as trapped ions, intra-beam scattering, instabilities etc., as well as complex beam dynamics [26,27], and the connection between damping time and injection rate is not well established. Table 2 gives data for three storage rings with a low energy injection, showing that multi-cycle injection can be achieved with a period as short as 1/25th of a damping time. At even lower energies a multi-cycle injection becomes impossible, however, it may be possible to inject sufficient current in a single shot. For example, the 600 MeV Super-ALIS ring in Japan can be injected in this way at only 15 MeV, where the radiation damping time is very long indeed ( $\sim 4$  min) [28].

## 6 . MODIFICATION OF DAMPING RATES

In the following sections we consider various ways in which the damping rates can be modified in an existing lattice.

### 6.1 Gradient wiggler magnet

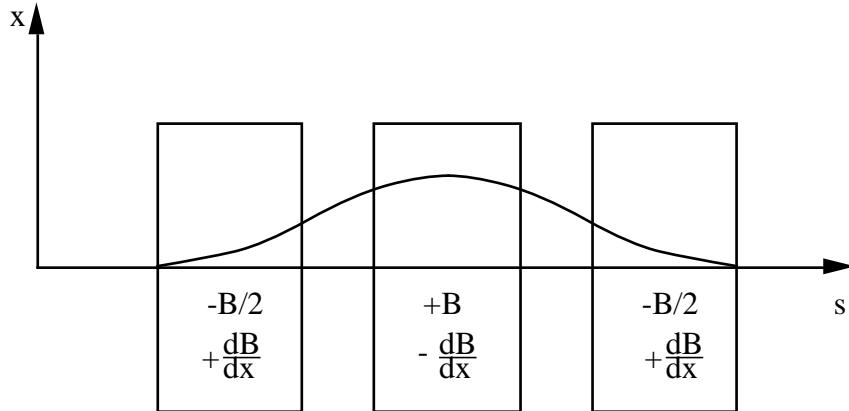


Fig. 8 Schematic diagram of a gradient wiggler magnet.

In order to modify the damping partition between the three planes a wiggler magnet with a gradient field may be used. This was first proposed by Robinson in 1958 as a means of overcoming the radial anti-damping of the CEA electron synchrotron, which has a combined function lattice [8,17]. For this reason it is often referred to as a "Robinson wiggler".

The method is to reduce the damping of the energy oscillation, thereby increasing the damping of the radial motion, by using a magnet in which higher energy electrons radiate less than lower energy electrons i.e.  $dU/d\epsilon$  is reduced. From Eq.(10) it can be seen that  $D$  reduces if  $2kD/\rho > 0$  i.e.  $DB(dB/dx) < 0$ . A series of magnet poles with alternating polarity of dipole and gradient fields as shown in Fig. 8 will therefore achieve this. Such magnets were installed at CEA in order to permit operation as a storage ring [25,31] and also in the PS to permit operation with electrons [5]. The magnets used in the latter case are shown in Fig. 9.

Gradient wigglers have also been proposed as a means of decreasing the beam emittance in various synchrotron radiation sources, as will be discussed in the following Chapter.

### 6.2 Variation of r.f. frequency

Another technique that can be employed in large rings for modifying the damping partition numbers is variation of the r.f. frequency [20]. The effect of a change in frequency ( $f$ ) is to cause the orbit length ( $L$ ) to vary, so forcing the electrons to move onto an off-energy orbit:

$$\frac{\Delta f}{f} = -\frac{\Delta L}{L} = -\alpha \frac{\epsilon}{E_0}$$

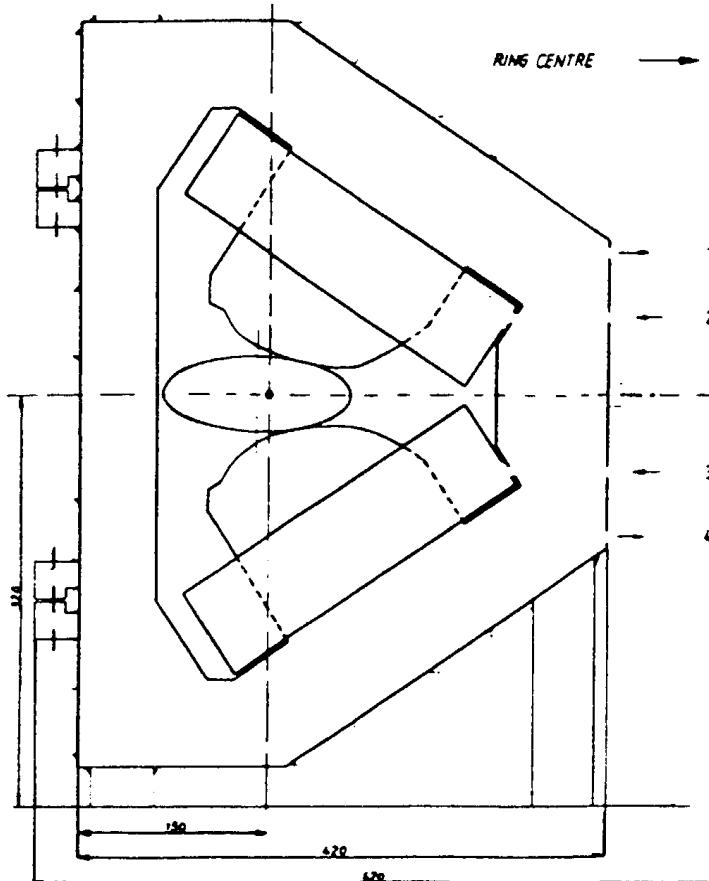


Fig. 9 Cross-section of the gradient wiggler magnet used in the PS [5]

where  $\alpha$  is the momentum compaction factor. The shift in the orbit where there is finite dispersion,  $x_\epsilon = D(s) \epsilon/E_0$ , has various effects, the largest of which results from the dipole field seen by the particle in the quadrupole magnets,  $(1/\rho)_{\text{quads}} = -kD\epsilon/E_0$  [32]. It follows from Eq. (10) that there is a change in  $D$  given by:

$$\Delta D \equiv \frac{\oint 2D^2 k^2 \epsilon / E_0 ds}{\oint 1/\rho^2 ds}$$

Table 3 shows the magnitude of the effect for three rings of different size, expressed as the change in the horizontal damping partition number with energy deviation and with mean orbit radius ( $R = L/2\pi$ ). It can be seen that in order to change  $J_x$  by unity the mean orbit position needs to be shifted by only 0.5 mm in LEP, with a corresponding energy deviation of only 0.13%, whereas in the EPA this change would require a movement of 30 mm, with an energy deviation of 5.6%. The method is effective therefore only in large rings; for example it was used regularly in PETRA [33] for luminosity optimization.

**Table 3**

Variation of horizontal damping partition number in various electron storage rings

$dJ_x/d(\epsilon/E_0)$	EPA [5]	PEP [34]	LEP [6]
$dJ_x/dR$ (mm $^{-1}$ )	-18 -0.03	-100 -0.26	-764 -0.47

### 6.3 Betatron coupling

The techniques discussed in the above two sections both involve changing the damping partition between the horizontal betatron motion and the energy oscillations. Another possibility is to vary the partition between the horizontal and vertical planes by means of skew-quadrupole or solenoidal fields, without affecting the energy oscillations. This was one of the additional techniques originally proposed as a means of overcoming the radial anti-damping in combined function lattices [8,20].

### 6.4 Dipole wiggler magnet

It follows from Eq. (16) that an increase in the energy loss per turn  $U_0$  will bring about an increase in all three damping rates. This can be achieved using a series of magnets with alternating polarity, arranged so that there is no net deflection of the electron beam as shown in Fig. 8, but in this case without the gradient field. Such a device is known as a dipole wiggler, or alternatively as a damping wiggler. A dipole wiggler also affects the equilibrium between the radiation damping and quantum excitation processes and so modifies the emittance in a complicated way, depending on the ring energy, wiggler parameters and the dispersion function, and will be discussed further in the following Chapter. Dipole wigglers are in operation in LEP and Fig. 10 shows the design of the magnets that are used [35].

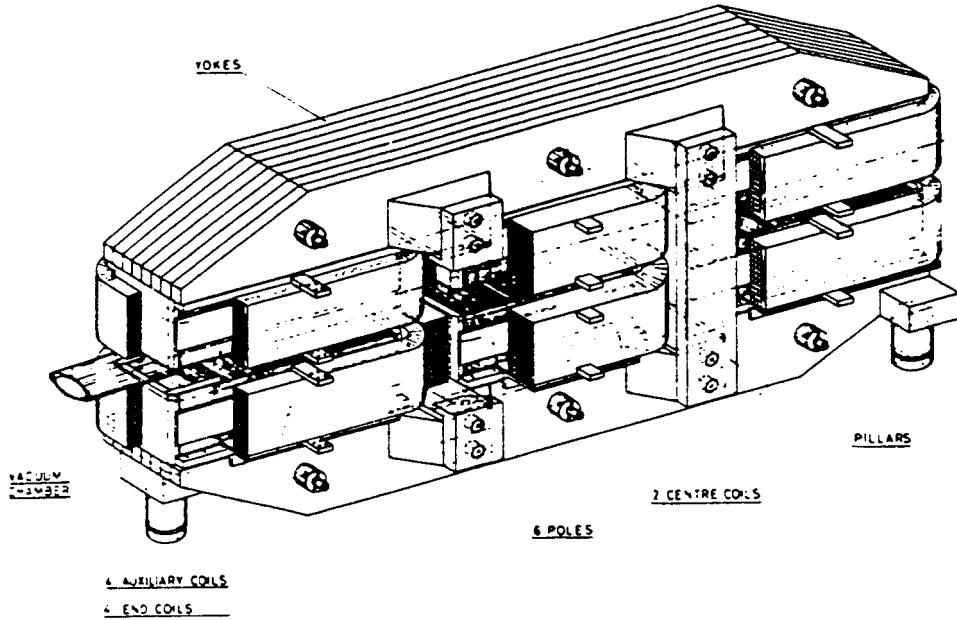


Fig. 10 The dipole wiggler magnet design for LEP [35]

## 7. MEASUREMENT OF DAMPING RATES

In general, the damping rates are of less interest as compared with other parameters, such as for example beam sizes and bunch lengths, and for this reason there are few published reports about such measurements. Several measurements have however been made at the SLC damping rings, whose performance depends very much on the damping rate. In one experiment the sum of the three damping rates was inferred indirectly by measuring the energy loss per turn  $U_0$  (Eq. (16)) [36]. This was done by using the relation  $U_0 = V_{rf} \sin (\phi)$ , by measuring the peak accelerating voltage ( $V_{rf}$ ) and the phase angle ( $\phi$ ), extrapolated to zero

current. The damping time was also obtained by measuring the variation of the extracted beam size as a function of storage time. More recently a synchrotron light monitor was used with a fast gated camera to directly measure the beam size as a function of time after injection. Figure 11 shows a typical result [37].

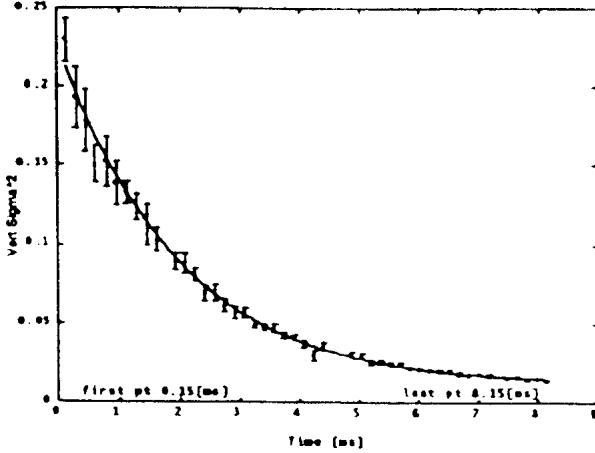


Fig. 11 Variation of the beam size in the SLC electron damping ring as a function of time after injection [37]

By fitting the data with an expression of the form:

$$\sigma_{\beta}^2 = \sigma_{\beta i}^2 \exp\left(-\frac{2t}{\tau}\right) + \sigma_{\beta 0}^2 \left[1 - \exp\left(-\frac{2t}{\tau}\right)\right],$$

where  $\sigma_{\beta 0}$  is the r.m.s. equilibrium beam size and  $\sigma_{\beta i}$  is the initial value after injection, the damping times in both horizontal and vertical planes ( $\tau_x$  and  $\tau_z$ ) were obtained.

A similar method was used also in the EPA to measure the horizontal and vertical damping rates. In this case a stored beam was excited with a fast kicker magnet and the changing beam profile observed with a synchrotron radiation beam profile monitor [38]. Figure 12 shows a sample result, from which the damping time may be extracted using the expression above, with  $\sigma_{\beta i}$  equal to the initial value after the blow-up using the kicker magnet (assuming zero dispersion at the measurement point).

For the longitudinal damping rate there is the possibility of making the same observations as above in the horizontal plane but at a point with large dispersion. This is because the total beam size contains contributions from the betatron motion and the energy spread,  $\sigma_{\text{total}}^2 = \sigma_{\beta}^2 + D^2(\sigma_e^2/E_0^2)$  with different damping rates,  $\tau_x$  and  $\tau_e$  respectively. Alternatively, some form of r.f. excitation could be applied and the resulting changes in the bunch length could be measured directly using an appropriate electron beam pick-up or synchrotron light monitor system.

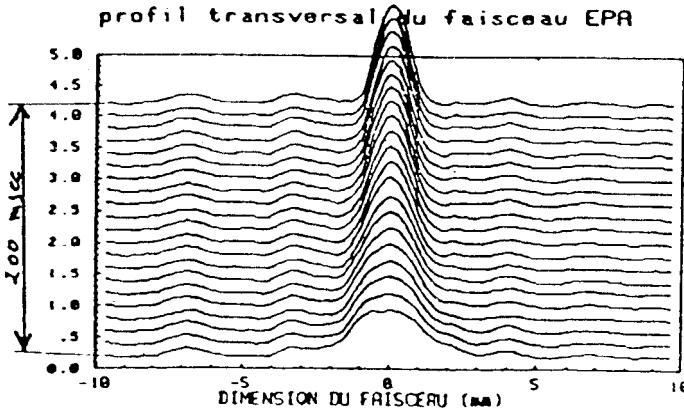


Fig. 12 Variation of the transverse profile of the beam in EPA after excitation with a fast kicker magnet [38]

## REFERENCES

- [1] M. Sands, SLAC-121 (1970).
- [2] S. Krinsky, M.L. Perlman and R.E. Watson, "Characteristics of Synchrotron Radiation and its Sources", Handbook on Synchrotron Radiation, Vol. Ia, North Holland, 1983.
- [3] K. Hübner, Proc. CERN Accelerator School, General Accelerator Physics, CERN 85-19, Nov. 1985, p. 239.
- [4] J. Le Duff, These proceedings.
- [5] LEP Design Report, Vol.1, The LEP Injector Chain, CERN-LEP/TH/83-29, June 1983.
- [6] LEP Design Report, Vol.2, The LEP Main Ring, CERN-LEP/84-01, June 1984.
- [7] J.M. Jowett, AIP Conference Proceedings 153, Vol.1, Physics of Particle Accelerators, American Inst. Phys. (1987) p. 864.
- [8] K.W. Robinson, Phys. Rev. 111 (1958) 373.
- [9] Advanced Engineering Mathematics, E. Kreyszig, John Wiley (1988).
- [10] K. Steffen, Ref. 3, p. 25.
- [11] N.H. Frank, Phys. Rev. 70 (1946) 177.
- [12] D. Bohm and L. Foldy, Phys. Rev. 70 (1946) 249.
- [13] J. Schwinger, Phys. Rev. 70 (1946) 798.
- [14] G. Rakowsky and L.R. Hughey, IEEE Trans. Nucl. Sci. NS-26 (1979) 3845.
- [15] H. Yamada, J. Vac. Sci. Technol. B8, (1990) 1628.
- [16] A.A. Kolomenski and A.N. Lebedev, CERN Symposium (1956) p. 447.
- [17] M.S. Livingston, CERN Symposium (1956) p. 439.

- [18] I.G. Henry, Phys. Rev. 106 (1957) 1057.
- [19] Yu.F. Orlov et. al., Proc. Int. Conf. High Energy Accelerators, CERN (1958) p. 306.
- [20] H.G. Hereward, Proc. Int. Conf. High Energy Accelerators, Brookhaven (1961) p. 222.
- [21] A. Hofmann and B. Zotter, IEEE Trans. Nucl. Sci. NS-24 (1977) 1875.
- [22] L. Mango, Frascati National Laboratory, LNF-62/34, May 1962.
- [23] F.E. Mills, Nucl. Instr. Meth. 23 (1963) 197.
- [24] G. Hemmie, IEEE Trans. Nucl. Sci. NS-30 (1983) 2028.
- [25] H. Winick, IEEE Trans. Nucl. Sci., NS-20 (1973) 984.
- [26] E. Weihreter, Proc. 3rd European Particle Accelerator Conference, Berlin, March 1992, Editions Frontières (1992), p. 93.
- [27] M. Eriksson, Nucl. Instr. Meth. Phys. Res. A291 (1990) 461.
- [28] T. Hosokawa et. al., Rev. Sci. Instr. 60 (1989) 1783.
- [29] E. Weireter et. al., Proc. 2nd European Particle Accelerator Conference, Nice, June 1990, Editions Frontières (1990), p. 237.
- [30] E.M. Rowe, Proc. 1987 US PArticle Accelerator Conference, IEEE CH2387-9/87/0000-0391, (1987) p. 391.
- [31] A. Hofmann et. al., Proc. 6th Int. Conf. High Energy Accelerators, Cambridge, 1967, p. 123.
- [32] J.P. Delahaye and J. Jäger, SLAC-PUB-3585, February 1985.
- [33] A. Piwinski, IEEE Trans. Nucl. Sci. NS-30 (1983) 2378.
- [34] M. Donald et al, SSRL-ACD-Note 42, December 1986.
- [35] J.M. Jowett and T.M. Taylor, IEEE Trans. Nucl. Sci. NS-30 (1983) 2581.
- [36] L. Rivkin et al, IEEE Trans. Nucl. Sci. NS-32 (1985) 2626.
- [37] F.J. Decker et. al., Proc. 3rd European Particle Accelerator Conference, Berlin, March 1992, Editions Frontières, p. 342.
- [38] J.F. Bottollier and J.P. Delahaye, CERN PS/LPI Note 87-11, March 1987.

# QUANTUM EXCITATION AND EQUILIBRIUM BEAM PROPERTIES

R.P. Walker  
Sincrotrone Trieste, Italy

## Abstract

Effects arising from the discrete, or quantized, nature of the emission of synchrotron radiation are considered. Combined with the results of the previous Chapter on radiation damping, the equilibrium beam distributions and loss of particles due to finite acceptance (quantum lifetime) are derived. The changes in the equilibrium values that are introduced by insertion devices are also considered.

## 1. INTRODUCTION

In the previous Chapter it was shown that the loss of energy due to the emission of synchrotron radiation (SR) and its replacement in the r.f. cavities can give rise to a *damping* of the betatron and synchrotron oscillations. When this process was first understood it led to speculation that the bunch size would eventually become so compressed that emission of coherent radiation would set a severe limit on the maximum beam intensity [1–3]. However it was later realized, firstly in connection with the energy oscillations [4,5] and later also for the betatron oscillations [6,7], that the emission of SR gives rise to another effect – quantum excitation – that causes a *growth* in the oscillation amplitudes, and that the combination of the two effects can give result in a stable equilibrium.

How can the emission of SR give rise to both a damping and excitation? As shown in the previous Chapter, radiation damping is related to the *continuous* loss and replacement of energy. However, energy is lost in discrete units or "quanta", i.e. photons, whose energy and time of emission vary randomly. This randomness introduces a type of noise or diffusion, causing growth of the oscillation amplitudes. The damping effect is linearly proportional to the energy loss or gain,  $\delta\epsilon$ , and so the total effect depends on the sum of such events,  $\Sigma (\delta\epsilon)_i$ , and therefore on the total energy loss per turn  $U_0$ , independent of how the photons are distributed in energy. However, as will be seen later, the quantum effect depends on  $(\delta\epsilon)^2$  and so the total effect is no longer simply related to the total energy loss per turn –  $\Sigma (\delta\epsilon)_i^2 \neq U_0^2$  – but depends on the numbers of photons with different photon energies i.e. the photon distribution function. A further distinguishing feature of expressions that describe the quantum excitation is that they all contain Planck's constant,  $h$ , whereas in the expressions for the radiation damping this factor is absent.

In this Chapter we consider the equilibrium distribution of the particles that results from the combined effect of quantum excitation and radiation damping, and derive expressions for the emittance, energy spread and bunch length. An estimate is also made of the rate of loss of particles resulting from the finite acceptance for the betatron and synchrotron oscillations, known as the quantum lifetime. A basic approach has been taken, following closely that of Sands [8]; more sophisticated treatments may be found in Refs. [9,10]. An introduction is also made to the topics of low emittance lattices and the effect on the equilibrium beam properties caused by insertion devices.

## 2. ENERGY OSCILLATIONS

### 2.1 Mean-square energy deviation

We recall the basic equations for the energy oscillations, with no damping:

$$\varepsilon(t) = A \cos(\Omega t - \phi) \quad (1)$$

$$\tau(t) = \frac{-\alpha}{E_0 \Omega} A \sin(\Omega t - \phi) \quad (2)$$

where  $\alpha$  is the momentum compaction factor and  $\Omega$  the synchrotron oscillation frequency. The invariate oscillation amplitude is thus given by:

$$A^2 = \varepsilon^2(t) + \left( \frac{E_0 \Omega}{\alpha} \right)^2 \tau^2(t). \quad (3)$$

When a photon is emitted the energy deviation changes,  $\varepsilon \rightarrow (\varepsilon - u)$ , and so the change in  $A^2$  is therefore:

$$\delta A^2 = -2\varepsilon u + u^2 \quad (4)$$

The first term is linear in  $u$  and corresponds to the radiation damping, as can be seen as follows. If the energy loss ( $u$ ) were independent of energy deviation ( $\varepsilon$ ) then over the synchrotron oscillation period this term would average to zero, i.e. no damping. However, by including the linear part of the variation of the energy loss with energy deviation, and averaging over one turn, we arrive at an equivalent expression for the damping as derived in the previous Chapter but in terms of  $A^2$  rather than  $A$ :

$$\frac{dA^2}{dt} = \frac{-A^2}{T_0} \frac{dU}{dE} = \frac{-2A^2}{\tau_\varepsilon} \quad (5)$$

where  $\tau_\varepsilon$  is the synchrotron oscillation damping time.

We will consider now the second term in the above, which being always positive can be seen to give rise to a growth in  $A^2$ . Since each emission is independent the average rate of increase is obtained by summing the effect of the  $n(u)$   $du$  photons emitted in each energy interval  $du$ :

$$\left\langle \frac{dA^2}{dt} \right\rangle = \int_0^\infty u^2 n(u) du = N \langle u^2 \rangle \quad (6)$$

where  $n(u)$  is the photon distribution function introduced in the Chapter on Synchrotron Radiation. Both  $N$  and  $\langle u^2 \rangle$  vary around the orbit, however since the effects that we are interested in occur slowly with respect to the orbit time we may average over many turns. Also, it can be shown that the betatron and synchrotron oscillations have only a small effect, and so we may simply take the average over the design orbit. Including the radiation damping term, we have therefore the following total rate of change of  $A^2$ :

$$\frac{dA^2}{dt} = \frac{-2A^2}{\tau_\varepsilon} + \left\langle N \langle u^2 \rangle \right\rangle \quad (7)$$

An equilibrium is reached when  $dA^2/dt = 0$ , in which case the mean value of  $A^2$  is given by:

$$\langle A^2 \rangle = \frac{\tau_\varepsilon}{2} \left\langle N \langle u^2 \rangle \right\rangle \quad (8)$$

It follows that the mean-square equilibrium value of the energy deviation is therefore:

$$\langle \epsilon^2 \rangle = \frac{\langle A^2 \rangle}{2} = \frac{\tau_\epsilon}{4} \langle N \langle u^2 \rangle \rangle \quad (9)$$

## 2.2 Distribution of the energy deviation

The above calculation results in a value of the mean-square energy deviation, but tells us nothing about the distribution function of the energy deviation, which is also of interest. In the approximation used so far that the energy oscillations are linear with respect to energy deviation, we can write an expression for the energy deviation at a given time as a sum of all the previous photon emissions, including the damping term:

$$\epsilon(t) = \sum_{i,t>t_i} u_i \exp\left[\frac{-(t-t_i)}{\tau_\epsilon}\right] \cos[\Omega(t-t_i)] \quad (10)$$

Since the typical energy deviation far exceeds the typical photon energy, the sum therefore contains a large number of small terms. These terms are also statistically independent and equally positive or negative (due to the phase factor). Therefore, according to the Central Limit Theorem of probability theory [11], the resulting distribution of the energy deviation is Gaussian, *independent* of the probability distribution function for  $u$ . Furthermore, the variance of the distribution is equal to the sum of the variances of the individual terms:

$$\langle \epsilon^2 \rangle = \sum_i \frac{\langle u^2 \rangle}{2} \exp\left[\frac{-2(t-t_i)}{\tau_\epsilon}\right] \quad (11)$$

Approximating as an integral and evaluating then gives:

$$\langle \epsilon^2 \rangle = \frac{N \langle u^2 \rangle}{2} \int_{-\infty}^t \exp\left[\frac{-2(t-t_i)}{\tau_\epsilon}\right] dt_i = \frac{N \langle u^2 \rangle \tau_\epsilon}{4} \quad (12)$$

in agreement with Eq. (9). Thus, we can relate the previous mean-square deviation to the standard deviation of the Gaussian distribution for the energy deviation:

$$\sigma_\epsilon^2 = \langle \epsilon^2 \rangle = \langle N \langle u^2 \rangle \rangle \frac{\tau_\epsilon}{4} \quad (13)$$

It is interesting to note that, as one might expect, the resulting value corresponds closely to the statistical uncertainty in the number of photon emissions that occur in one damping time, multiplied by the typical photon energy:

$$\sigma_\epsilon \equiv \sqrt{N \tau_\epsilon} u_c \quad (14)$$

## 2.3 Equilibrium energy spread

We use the following results from the previous Chapters:

$$\langle u_c^2 \rangle = \frac{11}{27} u_c^2 \quad N = \frac{15\sqrt{3}}{8} \frac{P}{u_c} \quad (15)$$

and therefore:

$$\langle N \langle u^2 \rangle \rangle = \frac{55}{24 \sqrt{3}} \langle P u_c \rangle \quad (16)$$

where,

$$P = \frac{e^2 c}{6\pi \epsilon_0} \frac{\gamma^4}{\rho^2} \quad u_c = \frac{3}{2} \frac{\hbar c \gamma^3}{\rho} \quad (17)$$

Also,

$$\tau_\varepsilon = \frac{2 E_0 T_0}{J_\varepsilon U_0} = \frac{2 E_0}{J_\varepsilon \langle P \rangle} \quad (18)$$

Inserting in Eq. (13) results in:

$$\sigma_\varepsilon^2 = \frac{55}{32 \sqrt{3}} \hbar c \gamma^3 \frac{\langle 1/\rho^3 \rangle}{\langle 1/\rho^2 \rangle} \frac{E_0}{J_\varepsilon} \quad (19)$$

The relative energy deviation is then:

$$\left( \frac{\sigma_\varepsilon}{E_0} \right)^2 = \frac{55}{32 \sqrt{3}} \frac{\hbar}{mc} \frac{\gamma^2}{J_\varepsilon} \frac{\langle 1/\rho^3 \rangle}{\langle 1/\rho^2 \rangle} \quad (20)$$

which simplifies in the isomagnetic case to:

$$\left( \frac{\sigma_\varepsilon}{E_0} \right)^2 = C_q \frac{\gamma^2}{J_\varepsilon \rho} \quad (21)$$

where the constant  $C_q$  is defined by:

$$C_q = \frac{55}{32 \sqrt{3}} \frac{\hbar}{mc} = 3.84 \cdot 10^{-13} \text{ m} \quad (\text{electrons}) \quad (22)$$

Since in most existing rings the chosen bending radius increases roughly as energy [2], the resulting energy spread values are very similar, typically about 0.1% as can be seen from the examples given in the table below.

**Table 1**  
Energy spread in various electron storage rings.

Ring	$E$ (GeV)	$\rho$ (m)	$\sigma_\varepsilon/E$ (%)
EPA	0.6	1.43	0.06
ESRF	6.0	25	0.10
PEP	18.0	166	0.12
LEP	55.0	3100	0.08

## 2.4 Equilibrium bunch length

A Gaussian distribution in energy results in a similar distribution in the time deviation  $\tau$ , and hence a Gaussian bunch shape in the longitudinal direction with standard deviation given by:

$$\sigma_\tau = \frac{\alpha}{\Omega E_0} \sigma_\varepsilon \quad (23)$$

In the isomagnetic case therefore:

$$\sigma_\tau^2 = \left( \frac{a}{\Omega} \right)^2 C_q \frac{\gamma^2}{J_\varepsilon \rho} \quad (24)$$

Inserting the expression for the synchrotron oscillation frequency,

$$\Omega^2 = \frac{\alpha}{T_0 E_0} eV_0 \quad (25)$$

we obtain:

$$\sigma_\tau^2 = \frac{C_q}{(mc^2)^2} \frac{\alpha T_0}{J_\varepsilon \rho} \frac{E_0^3}{eV_0} \quad (26)$$

Expressing the slope of the r.f. voltage in terms of the overvoltage,  $q = e\hat{V}/U_0$ , for the case of a sinusoidal variation in voltage with peak value  $\hat{V}$ :

$$eV_0 = (q^2 - 1)^{1/2} U_0 \omega_{r.f.} \quad (27)$$

gives the following:

$$\sigma_\tau^2 = \frac{E_1}{2\pi} \frac{\alpha T_0}{J_\varepsilon E_0} \frac{1}{(q^2 - 1)^{1/2}} \frac{1}{\omega_{r.f.}} \quad (28)$$

where  $\omega_{r.f.}$  is the angular r.f. frequency,  $E_1$  is a constant  $= (55\sqrt{3}/64) \hbar c / r_0 = 1.042 \cdot 10^8$  eV, and  $r_0$  is the classical electron radius ( $2.818 \cdot 10^{-15}$  m).

The bunch length,  $c\sigma_\tau$ , thus depends on many parameters such as energy, r.f. frequency and voltage, and the momentum compaction factor, which depends on the lattice design. Typical bunch lengths lie in the range 1–5 cm, however there are wide variations as can be seen from the examples in the table below:

**Table 2**  
Natural bunch lengths in various electron storage rings

Ring	$E$ (GeV)	r.f. (MHz)	$c\sigma_\tau$ (cm)
EPA	0.6	19.3	25
SLC damping ring	1.2	714	0.5
ESRF	6.0	352	0.5
LEP	55.0	352	1.6

In a given ring the bunch length is most commonly adjusted by changing the r.f. voltage. If the overvoltage ( $q$ ) is large, then Eq. (28) shows that the bunch length is inversely proportional to  $\sqrt{q}$ . Another possibility is to change the r.f. frequency, by using a separate set of accelerating cavities. For example, using two r.f. systems (62.4 and 500 MHz) and r.f.

voltage adjustment the bunch length in the BESSY storage ring was varied over a *wIDe* range between 0.7 and 8 cm [12]. A further technique that has been used in some cases is an additional r.f. cavity operating on a higher harmonic of the r.f. frequency in order to change the slope of the r.f. voltage,  $V_0$  in Eq. (26).

It should be noted that the bunch length calculated above is usually only obtained in practice with very small beam currents. Most rings exhibit the phenomenon of "bunch lengthening" as a function of the beam current, due a collective interaction of the beam with its surroundings.

### 3. BETATRON OSCILLATIONS

#### 3.1 Horizontal plane

We recall from the previous Chapter that photon emission at a point with non-zero dispersion gives rise to a change in the off-energy orbit, and hence introduces a change in the betatron motion. For an individual photon of energy  $u$  therefore:

$$\delta x_\beta = -D(s) \frac{u}{E_0} \quad \delta x'_\beta = -D'(s) \frac{u}{E_0} \quad (29)$$

The betatron oscillation invariant is given by:

$$A^2 = \gamma x^2 + 2\alpha x x' + \beta x'^2 \quad (30)$$

and hence the change due to the photon emission is therefore:

$$\delta A^2 = (\gamma D^2 + 2\alpha D D' + \beta D'^2) \frac{u^2}{E_0^2} \quad (31)$$

Only terms in  $u^2$  have been included, since the linear terms correspond to the radiation damping, as was the case with the energy oscillations.

Defining the important quantity,  $H$ :

$$H(s) = \gamma D^2 + 2\alpha D D' + \beta D'^2 \quad (32)$$

and following the same procedure as for the energy oscillations, the average rate of increase of  $A^2$  is then given by:

$$\left\langle \frac{dA^2}{dt} \right\rangle = \frac{\left\langle N \langle u^2 \rangle H \right\rangle}{E_0^2} \quad (33)$$

where as before the average is taken around the design orbit. Including the radiation damping term:

$$\frac{dA^2}{dt} = \frac{-2A^2}{\tau_x} \quad (34)$$

results in an equilibrium with mean-square value given by:

$$\frac{\langle A^2 \rangle}{2} = \frac{\tau_x}{4} \frac{\langle N \langle u^2 \rangle H \rangle}{E_0^2} \quad (35)$$

This defines the important quantity known as the (horizontal) beam emittance,  $\epsilon_x$ , which by analogy with the earlier result for the energy oscillations is given as follows:

$$\epsilon_x = \frac{\langle A^2 \rangle}{2} = C_q \frac{\gamma^2}{J_x} \frac{\langle H / \rho^3 \rangle}{\langle 1 / \rho^2 \rangle} \quad (36)$$

The same argument about the cumulative effect of a large number of small deviations can be applied in this case also, leading to the conclusion that there is a Gaussian distribution in the conjugate variables  $x, x'$ . Figure 1 illustrates this distribution, which consists of a series of ellipses each with a constant value of  $A^2$ . The ellipse with  $A^2 = \epsilon_x$  defines the "1  $\sigma$ " contour and hence the r.m.s. beam size ( $\sigma_x$ ) and divergence ( $\sigma_{x'}$ ) of the distributions projected on the  $x, x'$  axes respectively.

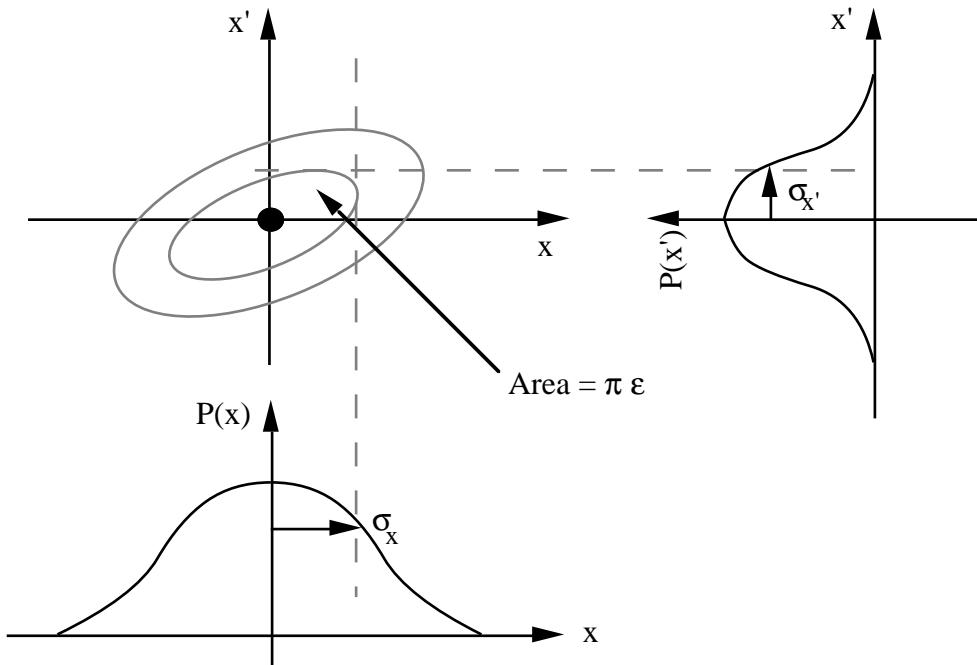


Fig. 1 Gaussian distribution of electron density in the  $(x, x')$  plane

It follows from the properties of the ellipse defined in Eq. (30) that:

$$\sigma_x = \sqrt{\epsilon \beta(s)} \quad \sigma_{x'} = \sqrt{\epsilon \gamma(s)} \quad (37)$$

Thus, although the emittance is a constant for a given lattice and energy, the beam size and divergence vary around the design orbit. At a symmetry point in the lattice therefore, where  $\alpha = 0$  and  $\gamma(s) = 1/\beta(s)$ , we have the simple interpretation that the emittance is the product of the beam size and divergence,  $\epsilon = \sigma_x \sigma_{x'}$ .

At a point in the lattice where there is finite dispersion the total horizontal beam size and divergence includes also a contribution from the energy spread. Since the betatron and synchrotron motions are uncorrelated the two widths add quadratically, and hence:

$$\sigma_x = \left\{ \varepsilon_x \beta_x(s) + D^2(s) \left( \frac{\sigma_\varepsilon}{E_0} \right)^2 \right\}^{1/2} \quad (38)$$

$$\sigma_x' = \left\{ \varepsilon_x \gamma_x(s) + D'^2(s) \left( \frac{\sigma_\varepsilon}{E_0} \right)^2 \right\}^{1/2}$$

In a given ring both the emittance and energy spread vary with  $E^2$ , and so the beam size and divergence vary linearly with energy. This is in sharp contrast to the situation for heavier particles for which radiation effects are negligible. In that case, the normalized phase-space area occupied by the beam is constant, in other words emittance is inversely proportional to energy.

### 3.2 Vertical plane

In the usual case of no bending and hence no dispersion in the vertical plane, the previous calculation would predict no quantum excitation and hence zero emittance in the vertical plane. A small effect arises due to the fact that the photons are not emitted exactly in the direction of the electron motion, which was neglected in the previous Section. If a photon of energy  $u$  is emitted at angle  $\theta_z$  with respect to the median plane, the change in angle of the electron is given by:

$$\delta z' = \frac{u}{E_0} \theta_z \quad (39)$$

and hence the change in vertical oscillation amplitude is:

$$\delta A^2 = \frac{u^2}{E_0^2} \theta_z^2 \beta_z(s) \quad (40)$$

By comparison with the previous formulae, and approximating as follows:

$$\langle u^2 \theta_z^2 \rangle \approx \langle u^2 \rangle \langle \theta_z^2 \rangle ; \langle \theta_z^2 \rangle \approx 1/2 \gamma^2 \quad (41)$$

the resulting equilibrium emittance becomes:

$$\varepsilon_z = \frac{C_q}{2} \frac{1}{J_z} \frac{\langle \beta_z / \rho^3 \rangle}{\langle 1/\rho^2 \rangle} \quad (42)$$

which in the isomagnetic case is:

$$\varepsilon_z = C_q \frac{\langle \beta_z \rangle}{2\rho} \quad \text{bending magnets} \quad (43)$$

where the average is taken over the bending magnets. Taking into account the value of  $C_q$ , Eq. (22), it can be seen that this value is very small indeed.

In practice the vertical emittance is not given by the value above, but arises from other processes:

- coupling of the horizontal and vertical betatron motion, arising from skew-quadrupole field errors. The latter can arise from angular positioning errors of the quadrupole magnets, and also from vertical closed orbit errors in sextupole magnets.
- vertical dispersion errors, arising from vertical bending fields produced by angular positioning errors of the dipoles, and vertical positioning errors of the quadrupoles.

The resulting vertical emittance thus depends only on errors, which can only be estimated statistically. It is common to describe the effect in terms of a coupling coefficient,  $\kappa$ , defined such that the sum of the horizontal and vertical emittances is constant:

$$\varepsilon_x = \frac{I}{I+\kappa} \varepsilon_{x_0} \quad \varepsilon_z = \frac{\kappa}{I+\kappa} \varepsilon_{x_0} \quad (44)$$

The quantity  $\varepsilon_{x_0}$ , calculated with Eq. (36), is often called the "natural beam emittance". The vertical beam size and divergence are then calculated as follows:

$$\sigma_z(s) = \sqrt{\varepsilon_z \beta_z(s)} \quad \sigma_z'(s) = \sqrt{\varepsilon_z \gamma_z(s)} \quad (45)$$

Typically, without correction, the coupling has a value of 1–10 %. In a given ring the coupling can be adjusted by means of an appropriate distribution of skew-quadrupole magnets, which excite a linear coupling resonance [13].

#### 4. SYNCHROTRON RADIATION INTEGRALS

The equations derived above and in the previous Chapter can be expressed in a general form that is valid also in the case of a non isomagnetic lattice, using the following Synchrotron Radiation Integrals [14]:

$$\begin{aligned} I_2 &= \oint \frac{1}{\rho^2} ds & I_3 &= \oint \frac{1}{|\rho^3|} ds \\ I_4 &= \oint \frac{D}{\rho} \left( \frac{1}{\rho^2} - 2k \right) = \oint \frac{(1-2n) D}{\rho^3} & I_5 &= \oint \frac{H}{|\rho^3|} ds \end{aligned} \quad (46)$$

It should be noted that a modulus sign has been included in some cases, in order that the correct values are obtained for elements with an opposite curvature to that of the main bending magnets e.g. in insertion devices. The beam parameters that can be calculated using the integrals are as follows:

Energy loss per turn:

$$U_0 = \frac{e^2}{6\pi \varepsilon_0} \gamma^4 I_2 \quad (47)$$

Damping partition numbers:

$$J_x = 1 - \frac{I_4}{I_2} \quad J_z = 1 \quad J_\epsilon = 2 + \frac{I_4}{I_2} \quad (48)$$

Damping times:

$$\tau_x = \frac{3 T_0}{r_0 \gamma^3} \frac{1}{I_2 - I_4} \quad \tau_z = \frac{3 T_0}{r_0 \gamma^3} \frac{1}{I_2} \quad t_\epsilon = \frac{3 T_0}{r_0 \gamma^3} \frac{1}{2I_2 + I_4} \quad (49)$$

Energy spread:

$$\left( \frac{\sigma_\epsilon}{E_0} \right)^2 = C_q \gamma^2 \frac{I_3}{2I_2 + I_4} = \frac{C_q \gamma^2}{J_\epsilon} \frac{I_3}{I_2} \quad (50)$$

Natural emittance:

$$\epsilon_{x_0} = C_q \gamma^2 \frac{I_5}{I_2 - I_4} = \frac{C_q \gamma^2}{J_x} \frac{I_5}{I_2} \quad (51)$$

where  $C_q$  is defined in Eq. (22). It can be noticed in the above that  $I_2$  and  $I_4$  are related to the radiation damping, whereas  $I_3$  and  $I_5$  are related to the quantum excitation.

At this point it is worth mentioning that the above equations are valid also for *protons*, with appropriate numerical values for  $r_0$  and  $C_q$ . In a given lattice therefore, the energy loss and damping times are reduced as the fourth power of the ratio of the masses of the particles ( $m_p/m_e = 1823$ ) and the energy spread and emittance as the third power. Thus, even in the case of the next generation of high energy proton machines such as LHC and SSC, the equilibrium values calculated from the above formulae are so small that in practice the beam dimensions will be limited by other processes such as intra-beam scattering.

## 5 . QUANTUM LIFETIME

### 5.1 Betatron oscillations

The distribution of beam intensity, both radially and vertically, is Gaussian and therefore in principle extends to infinity. However, the aperture defined by the vacuum chamber is finite and so there will be a constant loss of those electrons that approach the vacuum chamber walls. To calculate the effect we cannot simply use the probability distribution of the beam displacement ( $x$ ) at a given point in the ring, since the particles at a given  $x$  but lying on different phase ellipses will arrive eventually at different maximum values (see Fig. 1). We need therefore the probability distribution for the maximum value reached at that point in the ring,  $x_{\max}$ , or equivalently the invariant  $A^2$ , since  $x_{\max}^2 = A^2 \beta$ .

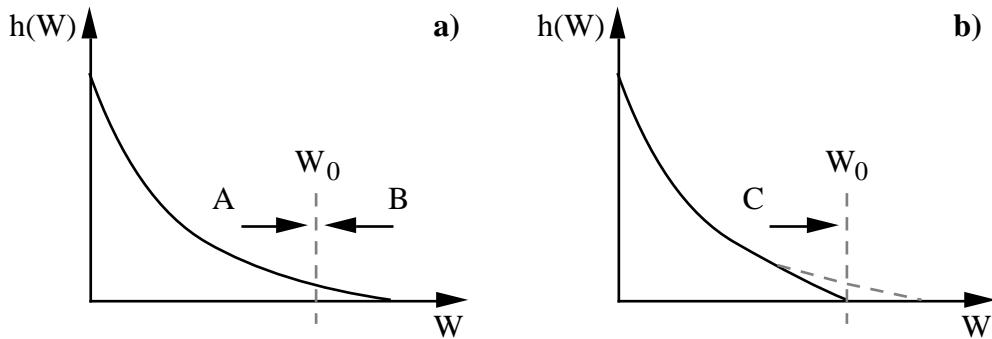


Fig. 2 Distribution of oscillation energies with no aperture (a) and with an aperture (b)

$A^2$ , which we shall call  $W$ , is a kind of "oscillation energy". It can be shown that the probability distribution for  $W$  is as follows:

$$h(W) = \frac{1}{\langle W \rangle} \exp \left( -\frac{W}{\langle W \rangle} \right) \quad (52)$$

Consider the number of electrons with oscillation energies increasing above or decreasing below a value  $W_0$  which corresponds to the aperture limit, i.e.  $W_0 = x_{\max}^2/\beta$ . Figure 2 illustrates the situation. With no aperture, and in the steady state, the number of electrons crossing  $W_0$  in each direction must be equal ( $A = B$ ). If the aperture is sufficiently far from the centre of the beam distribution the number of electrons which increase in amplitude, and hence are lost on the aperture limit, will be very nearly the same as if there were no aperture i.e.  $A \approx C$ . Hence the loss rate ( $C$ ) can be estimated by the rate at which particles cross the limit  $W_0$  due to radiation damping, in the steady state ( $B$ ). The rate is therefore:

$$\left( \frac{dN}{dt} \right)_{W_0} = \left( \frac{dN}{dW} \frac{dW}{dt} \right)_{W_0} \quad (53)$$

where:

$$\frac{dN}{dW} = N h(W) = \frac{N}{\langle W \rangle} \exp\left(\frac{-W_0}{\langle W \rangle}\right) \quad (54)$$

and:

$$\frac{dW}{dt} = \frac{-2W_0}{\tau_x}, \text{ since } W = \hat{W} \exp(-2t/\tau_x) \quad (55)$$

Hence:

$$\frac{dN}{dt} = -N \frac{2}{\tau_x} \frac{W_0}{\langle W \rangle} \exp\left(\frac{-W_0}{\langle W \rangle}\right) \quad (56)$$

The loss is therefore exponential:

$$N = N_0 \exp(-t/\tau_q); \frac{dN}{dt} = \frac{-N}{\tau_q} \quad (57)$$

where  $\tau_q$  is the "quantum lifetime" defined by:

$$\tau_q = \frac{\tau_x}{2} \frac{\langle W \rangle}{W_0} \exp\left(\frac{W_0}{\langle W \rangle}\right) \quad (58)$$

We can write this in the standard form:

$$\tau_q = \frac{\tau_x}{2} \frac{\exp(\xi)}{\xi} \quad (59)$$

where:

$$\xi = \frac{W_0}{\langle W \rangle} = \frac{x_{\max}^2}{2\sigma_x^2} \quad (60)$$

The lifetime is determined by the minimum value of  $x_{\max}/\sigma_x$ , or equivalently  $x_{\max}/\sqrt{\beta}$ , that occurs at some point around the ring (neglecting the effect of any closed orbit errors). This is known as the limiting acceptance, which because of the dependence on  $\beta$  may not correspond to the smallest physical aperture. Because of the exponential factor the quantum lifetime increases rapidly with  $\xi$  and hence  $x_{\max}/\sigma_x$ . To illustrate this fact, the table below gives values of the quantum lifetime for a range of values of  $x_{\max}/\sigma_x$  for a typical damping time of 10 ms.

**Table 3**  
Quantum lifetime as a function of limiting aperture.

$x_{\max}/\sigma_x$	5	5.5	6.0	6.5	7.0
$\tau_q$	1.8 min	20.4 min	5.1 h	98.3 h	103 days

Hence, one arrives at the 'Golden Rule' for long lifetime:

$$x_{\max}/\sigma_x \geq 6.5 \quad (61)$$

i.e. the beam aperture should be at least 13 times the r.m.s. beam size in order that the quantum lifetime does not play a significant part in determining the overall beam lifetime. Other processes will then dominate, such as scattering off residual gas molecules, Touschek scattering etc. [15].

## 5.2 Energy oscillations

There is also a quantum lifetime resulting from the finite r.f. acceptance for the energy oscillations. The synchrotron oscillations become non-linear at large energy deviations, but if we assume that the maximum possible energy deviation,  $\varepsilon_{\max}$ , is large compared to the r.m.s. energy deviation, the calculated loss rate should be approximately correct. We can estimate this loss rate in the same way as for the betatron oscillations. With the oscillation amplitude given by:

$$W = A^2 = \varepsilon^2 + \left( \frac{E_0 \Omega}{\alpha} \right)^2 \tau^2 \quad (62)$$

we obtain a similar result to Eq. (59):

$$\tau_q = \frac{\tau_\varepsilon}{2} \frac{\exp \xi}{\xi} \quad (63)$$

where:

$$\xi = \frac{W_0}{\langle W \rangle} = \frac{\varepsilon_{\max}^2}{2\sigma_\varepsilon^2} \quad (64)$$

Inserting expressions for  $\varepsilon_{\max}$  and  $\sigma_\varepsilon$  derived earlier we can write:

$$\xi = E_1 \frac{J_\varepsilon E_0}{\alpha h} F(q) \quad (65)$$

where  $F(q)$  is the 'energy aperture' function defined by [8]:

$$F(q) = 2 \left\{ \left( q^2 - 1 \right)^{1/2} - \cos^{-1} (1/q) \right\} \quad (66)$$

and  $h$  is the harmonic number of the r.f. system (r.f. frequency divided by orbit frequency).

The overvoltage ( $q$ ) required to ensure adequate quantum lifetime therefore is smallest in a lattice with small momentum compaction ( $\alpha$ ) and high harmonic number ( $h$ ). As an example, the 2 GeV SRS storage ring in its first phase had a relatively large momentum compaction ( $\alpha = 0.135$ ) which gave rise to a high overvoltage requirement of 7.2, for 100 hours quantum lifetime. After changing the magnet lattice in order to reduce beam emittance this also resulted in smaller momentum compaction ( $\alpha = 0.029$ ) and hence a significantly smaller overvoltage requirement of 2.7, with a consequent reduction in r.f. power demands.

It can be seen that similar terms appear in the expression above for  $\xi$  as in the expression for the bunch length, Eq. (28). We can use this fact to make a rough estimate for the bunch length under the conditions that the quantum lifetime is large. Combining the relevant equations we obtain:

$$\left( \frac{\sigma_t}{T_{\text{r.f.}}} \right)^2 = \frac{F(q)}{\xi (q^2 - 1)^{1/2}} \frac{1}{(2\pi)^2} \quad (67)$$

where  $T_{\text{r.f.}}$  is the r.f. period, i.e. the time interval between r.f. buckets. Approximating  $F(q)/(q^2 - 1)^{1/2}$  by its limiting value for large overvoltage (=2) then gives:

$$\left( \frac{\sigma_t}{T_{\text{r.f.}}} \right)^2 \approx \frac{1}{2\pi^2 \xi} \quad (68)$$

Thus, with a value of  $\xi = 21$  for good lifetime (equivalent to a value of  $\epsilon_{\text{max}}/\sigma_\epsilon = 6.5$ ), we obtain the simple result that the total bunch length (for example, the full wIDth half maximum) is about 10% of the bunch separation.

## 6 . LOW EMITTANCE LATTICES

Here we examine two applications of an electron storage ring in which a low emittance is a particular requirement.

### 6.1 Synchrotron radiation sources

A small beam size and divergence, i.e. a small beam emittance, is a general requirement of synchrotron radiation sources, in order to increase the brightness of the emitted radiation. From the expressions derived earlier, e.g. Eq. (51), it can be seen that to obtain low emittance requires a lattice design which minimizes the average H function in the bending magnets. In particular a small dispersion is required, which is not achieved in the classic separated function lattice, the FODO. Various types of lattice have therefore been developed in order to achieve this [16,17]. The first was the Chasman-Green (CG) structure, which is based on an achromatic arc composed of a pair of bending magnets with a focusing quadrupole in between [18]. Such a design results in zero dispersion in the straight sections between achromats which are therefore suitable locations for insertion devices (see Section 7). The limited flexibility of this lattice has led to the extended CG or double-bend achromat (DBA) and triple-bend achromat (TBA). For each lattice type there is a minimum achievable emittance which is given by an expression of the form:

$$\varepsilon_{x_0, \text{ min}} = f \frac{C_q \gamma^2}{J_x} \theta_b^3 \quad (69)$$

where  $\theta_b$  is the bending angle, assumed equal for all magnets. The factor  $f$  varies depending on the lattice type from 0.05 in the case of a DBA lattice to 0.36 for a FODO lattice. It should be noted however that in all cases the emittance increases as the square of the energy and varies inversely with the third-power of the number of bending magnets. In the case of several of the third generation synchrotron radiation sources that are under construction, the typical natural emittance is  $7 \cdot 10^{-9}$  m rad even though the rings vary widely in energy from 1.5 to 8 GeV. This is achieved by adjusting the number of achromats (and hence  $\theta_b$ ) between 10 and 44.

It can be seen from the equations in Section 4 that some reduction in the emittance can be obtained by increasing  $J_x$  which can be achieved by adding a vertically focusing field gradient ( $k$  and  $n$  positive) in the dipole magnets [19]. Some new storage rings (e.g. ALS, Berkeley, and ELETTRA, Trieste) employ such a gradient field both for emittance reduction and for optimization of the lattice  $\beta$  functions. In the latter case for example, the bending magnet has a field index of 13, giving  $J_x = 1.3$ .

## 6.2 Damping rings

A damping ring serves as a temporary storage ring to reduce the emittance of an injected beam by means of radiation damping. It can be seen from Eqs. (33) and (34) that a combination of quantum excitation and radiation damping processes leads to a general equation for the emittance, of the form:

$$\frac{d\varepsilon}{dt} = \text{constant} \cdot \frac{2\varepsilon}{\tau_x} \quad (70)$$

The emittance ( $\varepsilon$ ) therefore varies in time as follows:

$$\varepsilon(t) = \varepsilon_i \exp\left(-\frac{2t}{\tau_x}\right) + \varepsilon_0 \left[1 - \exp\left(-\frac{2t}{\tau_x}\right)\right] \quad (71)$$

where  $\varepsilon_0$  is the equilibrium emittance ( $t = \infty$ ) and  $\varepsilon_i$  is the injected beam emittance ( $t = 0$ ). For a given storage time, the optimum ring energy is thus a compromise between the need for both small equilibrium emittance (low energy) and fast damping (high energy). Since from Eq. (49) the damping time is proportional to  $T_0 \rho / E_0^3$  it follows that a fast damping requires also a small orbit circumference and small bending radius (high field strength). An example is the 1.21 GeV SLC positron damping ring [20], which has a small circumference (35 m) and bending radius (2 m, corresponding to a 2 T magnetic field) resulting in a small damping time of 3.1 ms, sufficient to reduce the initial positron beam emittance by about a factor of 300 with a storage time of 11.1 ms.

## 7. CHANGES IN BEAM PROPERTIES DUE TO INSERTION DEVICES

An insertion device (ID) is a magnetic device located in a straight section of a ring that produces a transverse field component that alternates in polarity along the beam direction. Such devices are used both as special sources of synchrotron radiation [see Chapter on Synchrotron Radiation] and as a means of controlling various beam parameters [21]. In general, IDs give rise to both additional radiation damping and quantum excitation, and so result in different equilibrium values of damping times, emittance and energy spread etc. which depend on the ID parameters and on the lattice functions at the ID location.

The Robinson, or gradient, wiggler introduced in the previous Chapter was the first type of insertion device, and was developed as a means of overcoming the radial anti-damping of combined function lattices. Such a device can also be used in separated function lattices that are already damped in all 3 planes as a means of reducing beam emittance. The dominating effect is the change introduced in the  $I_4$  integral, which affects the damping partition numbers. It can be seen directly from Eq. (51) that an increase in  $J_x$  from its usual value of 1 to 2 can reduce the emittance by a factor of 2, while still allowing damping of all oscillation modes.

The more common type of insertion device is the dipole or damping wiggler, which in general contributes to all of the Synchrotron radiation integrals. It can be seen from the equations in Section 4 that  $I_2$  always increases and hence the damping times all reduce, as described in the previous Chapter. The effects on energy spread and emittance are however more complicated. It follows from Eq. (50) that the ratio of the modified to the original (no ID) equilibrium values can in general be written as follows:

$$\frac{\sigma'_e}{\sigma_e} = \left[ \frac{1 + I_3^{ID}/I_3}{I + (2I_2^{ID} + I_4^{ID})/(2I_2 + I_4)} \right]^{1/2} \quad (72)$$

where the contributions of the insertion device to the integrals are labelled *ID*. In the common case of a sinusoidal field variation these contributions can be written as follows:

$$I_2^{ID} = \frac{L}{2\rho_{ID}^2} \quad I_3^{ID} = \frac{4}{3\pi} \frac{L}{\rho_{ID}^3} \quad I_4^{ID} = -\frac{1}{32\pi^2} \frac{\lambda_o^2}{\rho_{ID}^4} L \quad (73)$$

where  $L$  is the length of the ID,  $\lambda_o$  the period length and  $\rho_{ID}$  is the bending radius corresponding to the peak field of the ID. The  $I_4^{ID}$  term arises from the dispersion generated by the device itself, the so called self-dispersion. In most cases however it is negligible compared to the larger  $I_2^{ID}$  term. Simplifying for the isomagnetic lattice case, and also neglecting the  $I_4$  term, i.e. assuming  $J_x = 1$ , results finally in the following:

$$\frac{\sigma'_e}{\sigma_e} = \left[ \frac{1 + \frac{4}{3\pi} \frac{L}{2\pi\rho} \left( \frac{\rho}{\rho_{ID}} \right)^3}{1 + \frac{1}{2} \frac{L}{2\pi\rho} \left( \frac{\rho}{\rho_{ID}} \right)^2} \right]^{1/2} \quad (74)$$

It can be seen from the above that if the peak field in the ID is less than that of the bending magnets ( $\rho_{ID} > \rho$ ) there is a *reduction* in energy spread, whereas if the peak ID field exceeds the bending magnet field ( $\rho_{ID} < \rho$ ) the energy spread is *increased*.

In the case of the emittance the effect is complicated by the fact that the ID self-dispersion must be added to the dispersion that is present in the straight section without the ID:

$$D(s) = \left( \frac{\lambda_0}{2\pi} \right)^2 \frac{1}{\rho_{ID}} \cos\left( \frac{2\pi s}{\lambda_0} \right) \quad D'(s) = \left( \frac{\lambda_0}{2\pi} \right) \frac{1}{\rho_{ID}} \sin\left( \frac{2\pi s}{\lambda_0} \right) \quad (75)$$

The net result must in general be evaluated numerically, however some limiting cases can be examined. In the case that the dispersion in the straight section is large we can write (in the isomagnetic case):

$$\frac{\varepsilon_{x_0}'}{\varepsilon_{x_0}} = \left[ \frac{1 + \frac{4}{3\pi} \frac{L}{2\pi\rho} \left( \frac{H_{ID}}{H} \right) \left( \frac{\rho}{\rho_{ID}} \right)^3}{1 + \frac{1}{2} \frac{L}{2\pi\rho} \left( \frac{\rho}{\rho_{ID}} \right)^2} \right] \quad (76)$$

Thus the emittance can be increased or decreased depending on the relative values of  $H$  and  $\rho$  in the insertion device and in the bending magnets. It can be seen that if  $H/\rho$  in the ID exceeds  $H/\rho$  in the bending magnets then the emittance will be increased. In modern synchrotron radiation sources therefore, where it is usually wanted to preserve the low emittance, the insertion devices are usually placed in straight sections with zero dispersion. If however they are placed in a straight section with finite dispersion and if it is wanted to minimize the emittance increase then it can be seen from the above that the quantity  $\int 1/\rho' ds$  should be minimized in the magnet design.

In the case where there is zero dispersion in the straight section, the self-generated dispersion in the ID dominates. The largest term involved is:

$$I_5^{ID} = \frac{\lambda_0^2}{15\pi^3 \rho_{ID}^5} \langle \beta_x \rangle L \quad (77)$$

A similar expression can be derived in the case of a rectangular, rather than sinusoidal, field model. From the above a condition for the emittance not to be increased can be derived as follows:

$$\lambda_0^2 B^3 \leq 5.87 \cdot 10^9 \frac{E [\text{GeV}] \varepsilon_{x_0}}{\langle \beta_x \rangle} \quad (78)$$

It can be seen therefore that except for very high field devices in low emittance and low energy rings the emittance is generally reduced by the ID.

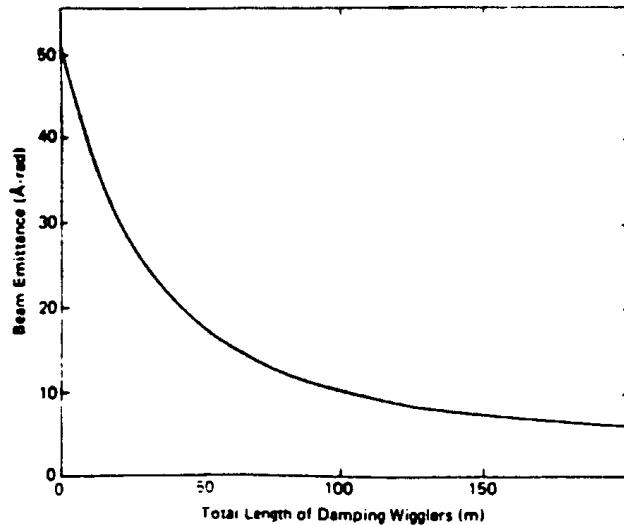


Fig. 3 Beam emittance in PEP as a function of the total damping wiggler length

It follows from the above that dipole wigglers may be used as a means of either increasing or decreasing the beam emittance. For example, they have been proposed as a means of obtaining a very low emittance in PEP, for operation as a synchrotron radiation source [22]. With a 1.26 T wiggler with a period length of 12 cm, Fig. 3 shows that a

reduction in emittance of nearly a factor of 10 can be achieved, albeit with a total wiggler length of some 200 m. The possibility of including dipole wigglers in the design of damping rings for the next generation of linear colliders has also been considered [23].

Two sets of dipole wigglers are in routine operation in LEP [24]. One set is located in a dispersion free region and is used at injection to increase the energy spread and bunch length by 5–6 times the normal value in order to improve beam stability. A second set is in a finite dispersion region and is used to increase the emittance in order to optimize the beam luminosity.

## REFERENCES

- [1] E.M. McMillan. Phys. Rev. 68 (1945) 144.
- [2] L.I. Schiff, Rev. Sci. Instr. 17 (1946) 6.
- [3] J.S. Nodvick and D.S. Saxon, Phys. Rev. 96 (1954) 180.
- [4] A.A. Sokolov and J.M. Ternov, Soviet Physics JETP 1 (1955) 227.
- [5] M. Sands, Phys. Rev. 97 (1955) 470.
- [6] A.A. Kolomenski and A.N. Lebedev, Proc. CERN Symposium on High Energy Accelerators, CERN, Geneva (1956) 447.
- [7] I.G. Henry, Phys. Rev. 106 (1957) 1057.
- [8] M. Sands, SLAC Report No. 121, Nov. 1970.
- [9] A. Piwinski, Proc. CERN Accelerator School, General Accelerator Physics, Paris, 1984, CERN 85-19, p. 432.
- [10] J. Jowett, Proc. CERN Accelerator School, Advanced Accelerator Physics, Oxford, 1985, CERN 87-03, p. 570.
- [11] H. Bruck, Nucl. Instr. Meth. 29 (1964) 54.
- [12] W. Anders, Proc. 3rd European PArticle Accelerator Conference, Berlin, March 1992, Editions Frontieres (1992) p. 798.
- [13] G. Guignard, Proc. CERN Accelerator School, Advanced Accelerator Physics, Oxford, 1985, CERN 87-03, p. 203.
- [14] R.H. Helm et. al., IEEE Trans. Nucl. Sci. NS-20 (1973) 900.
- [15] A. Wrulich, these Proceedings.
- [16] A. Ropert, Proc. CERN Accelerator School, Synchrotron Radiation and Free Electron Lasers, Chester, 1989, CERN 90-03, p. 158.
- [17] V.P. Suller, Proc. 3rd European PArticle Accelerator Conference, Berlin, March 1992, Editions Frontieres (1992) p. 77.
- [18] R. Chasman et. al., IEEE Trans. Nucl. Sci. NS-22 (1975) 1765.

- [19] G. Vignola, Nucl. Instr. Meth. Phys. Res. A246 (1986) 12.
- [20] J.P. Delahaye and L. Rivkin, IEEE Trans. Nucl. Sci. NS-32 (1985) 1695.
- [21] J.M. Paterson, SPEAR-186, July 1975.
- [22] H. Wiedemann, Nucl. Instr. Meth. Phys. Res. A266 (1988) 24.
- [23] T.O. Raubenheimer et. al., Proc. 1989 US particle Accelerator Conference, IEEE 89CH2669-0, p. 1316.
- [24] A. Hofmann, private communication (1992).